GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:37 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-5
Perfect score: 88
Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	999	1	R30742	Human pemphigus vu
2	88	100.0	15	1	W04845	Self epitope of de
3	83	94.3	614	1	W07908	Pemphigus vulgaris
4	74	84.1	778	1	W15489	Pemphigus foliaceu
5	60	68.2	787	1	R86865	Human protocadheri
6	52	59.1	832	1	W74089	Human HPT-1 protei
7	50	56.8	43	1	R58865	Rat-224 cadherin p
8	50	56.8	43	1	R87107	Protocadherin clon
9	50	56.8	1180	1	W61273	Homo sapiens proto
10	48	54.5	41	1	R58875	Rat-352 cadherin p
11	48	54.5	41	1	R87117	Protocadherin clon
12	48	54.5	780	1	W13130	Full length human
13	48	54.5	780	1	W25634	Human cadherin-5.
14	46	52.3	799	1	W13126	Full length rat ca
15	46	52.3	793	1	W13132	Full length human
16	46	52.3	532	1	W13127	Truncated rat cadh
17	46	52.3	793	1	W25635	Human cadherin-8.
18	46	52.3	799	1	W25639	Rat full length ca
19	46	52.3	532	1	W25640	Rat truncated cadh
20	46	52.3	560	1	W13009	Segment of desmoso
21	45	51.1	836	1	R58912	Product of alterna
22	45	51.1	682	1	R49144	Product of alterna
23	45	51.1	904	1	R58907	Human protocadheri
24	45	51.1	43	1	R58882	Mouse-321 cadherin
25	45	51.1	43	1	R58886	Human-11 cadherin-
26	45	51.1	148	1	R58876	Rat-411 cadherin p
27	45	51.1	40	1	R58878	Rat-551 cadherin p
28	45	51.1	43	1	R58868	Rat-314 cadherin p
29	45	51.1	682	1	R87154	Alternatively spli
30	45	51.1	836	1	R87153	Alternatively spli
31	45	51.1	904	1	R87147	Protocadherin clon
32	45	51.1	43	1	R87121	Protocadherin clon
33	45	51.1	41	1	R87120	Protocadherin clon
34	45	51.1	148	1	R87118	Protocadherin clon
35	45	51.1	43	1	R87125	Protocadherin clon
36	45	51.1	43	1	R87110	Protocadherin clon
37	45	51.1	794	1	W13135	Putative human cad
38	45	51.1	794	1	W25637	Human cadherin-12.
39	45	51.1	896	1	W41252	Xenopus paraxial p
40	44	50.0	391	1	R43564	Human cadherin-6.
41	44	50.0	1203	1	R58911	Product of alterna
42	44	50.0	1026	1	R58906	Human protocadheri
43	44	50.0	1203	1	R87152	Alternatively spli
44	44	50.0	797	1	R86866	Human protocadheri
45	44	50.0	1026	1	R87146	Protocadherin clon

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 88; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
|||||||||||||||
Db 251 CECNIKVKDVNDNFP 265

RESULT 2
W04845
ID W04845 standard; peptide; 15 AA.
AC W04845;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN WO9627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1; Page 40; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise

CC either an isolated human non-collagen or non-myslin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 251-265)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 88; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
||||||||||||||||
Db 1 CECNIKVKDVNDNFP 15

RESULT 3
W07908
ID W07908 standard; protein; 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphis vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucuous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 94.3%; Score 83; DB 1; Length 614;
Best Local Similarity 93.3%; Pred. No. 6.4e-06;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
||||||||| |||||
Db 250 CECNIKVKDYNDNFP 264

RESULT 4
W15489
ID W15489 standard; Protein; 778 AA.
AC W15489;
DT 17-JUN-1997 (first entry)
DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers
FT domain 1. .545
FT /note= "Pemphigus foliaceus antigen protein"
PN J09077800-A.
PD 25-MAR-1997.
PF 12-SEP-1995; 260899.
PR 12-SEP-1995; JP-260899.
PA (NISH/) NISHIKAWA T.
DR WPI; 97-241758/22.
DR P-PSDB; T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT through the hinge region used to treat pemphigus foliaceus
PS Claim 1; Page 10-12; 17pp; Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adsorbent upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is useful
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA;

Query Match 84.1%; Score 74; DB 1; Length 778;
Best Local Similarity 80.0%; Pred. No. 0.00025;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
|||||| |||| |
Db 253 CECNIKILDVNDNIP 267

RESULT 5
R86865
ID R86865 standard; Protein; 787 AA.
AC R86865;
DT 27-AUG-1996 (first entry)
DE Human protocadherin pc3.
KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW catenin; therapy.
OS Homo sapiens.
PN WO9600289-A1.
PD 04-JAN-1996.
PF 26-JUN-1995; U08071.
PR 27-JUN-1994; US-268161.
PA (DOHE-) DOHENY EYE INST.
PI Suzuki S;
DR WPI; 96-068873/07.
DR N-PSDB; T03572.
PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT pc5 - involved in cell-cell adhesion and regulation activities
PS Claim 15; Page 115-119; 146pp; English.
CC R86865-R86867 represent the sequences for three protocadherins. This
CC sequence represents the human protocadherin pc3. These sequences are
CC related to cadherin, and possess cell adhesive ability. Cadherins are
CC glycosylated integral membrane proteins that are involved in cell-cell
CC adhesion. Cadherins are composed of an N-terminal extracellular domain
CC which consists of 5 unique subdomains, a membrane spanning domain, and a
CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
CC cytoskeleton through catenins and other cytoskeleton associated proteins.
CC The cytoplasmic domain is not present in all cadherins, but in those
CC which possess it, it is essential for the cadherins adhesive function.
CC The cadherins which do not possess a cytoplasmic domain appear to
CC function via a different method from those with a cytoplasmic domain.
CC These sequences were isolated using primers 1 and 2 (see T03575 and
CC T03576) The proteins may have regulatory functions in the cell, as well
CC as the cell-cell adhesive properties. Antibodies produced against these
CC sequences are useful for modulating the binding activity of these
CC protocadherins, and can be used therapeutically.

```
SQ   Sequence   787 AA;

  Query Match             68.2%;  Score 60;  DB 1;  Length 787;
  Best Local Similarity   66.7%;  Pred. No. 0.049;
  Matches   10;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |  ::|| |||||||
Db     331 CSVSVKVLDVNDNFP 345


RESULT   6
W74089
ID   W74089 standard; Protein; 832 AA.
AC   W74089;
DT   04-MAY-1999  (first entry)
DE   Human HPT-1 protein sequence.
KW   Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW   D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW   intestinal peptide-associated transporter; hypertension; diabetes;
KW   osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW   therapeutic agent delivery; therapy; probe.
OS   Homo sapiens.
PN   WO9851325-A2.
PD   19-NOV-1998.
PF   15-MAY-1998; U10088.
PR   15-MAY-1997; US-046595.
PA   (CYTO-) CYTOGEN CORP.
PA   (ELAN-) ELAN CORP PLC.
PI   Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ,
PI   Omahony DJ, Patterson CA, Singleton J;
DR   WPI; 99-009568/01.
DR   N-PSDB; X18166.
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and
PT   antibodies, used to deliver therapeutic or diagnostic agents to, or
PT   through, the gastrointestinal tract, e.g. insulin or leuprolide
PS   Disclosure; Fig 2; 294pp; English.
CC   This sequence is the human HPT-1 protein. The invention relates to
CC   purified proteins (I) that bind specifically to at least one of the
CC   gastro-intestinal (GI) tract receptors human intestinal
CC   peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC   sucrose-isomaltase complex (hSI). (I) provide active transport of
CC   therapeutic agents through human and animal GI tissue (into the blood)
CC   for in vivo delivery, particularly for treatment or prevention
CC   of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC   migraine, or angina pectoris. Specifically they are used to deliver
CC   insulin or leuprolide, but many other suitable therapeutic agents are
CC   disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC   antigens. (I) may also provide targeting to the GI tract. Other uses of
CC   (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
SQ   Sequence   832 AA;

  Query Match             59.1%;  Score 52;  DB 1;  Length 832;
  Best Local Similarity   64.3%;  Pred. No. 1.1;
  Matches    9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ECNIKVKDVNDNFP 15
           | ::|||||:||| |
Db     325 EIHVKVKDINDNPP 338


RESULT   7
R58865
ID   R58865 standard; Protein; 43 AA.
AC   R58865;
DT   17-APR-1995  (first entry)
DE   Rat-224 cadherin partial sequence.
KW   Cadherin; cell adhesion molecule.
OS   Rattus rattus.
PN   WO9414960-A.
PD   07-JUL-1994.
PF   23-DEC-1993; U12588.
PR   29-DEC-1992; US-998003.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 94-293849/36.
DR   N-PSDB; Q68957.
PT   Polynucleotide sequences encoding new proto:cadherins - useful
PT   for modulating natural binding and regulating activities.
PS   Example; Page 38; 114pp; English.
CC   Two regions of conserved AA sequence, one from the middle of the
CC   third cadherin extracellular subdomain (EC-3) and the other from the
CC   C-terminus of the fourth extracellular subdomain (EC-4) were
CC   identified. The corresp. degenerate oligos (Q68949, Q68950) were
CC   designed for use as PCR primers. PCR was carried out on a rat brain
CC   cDNA prepn. Two major bands of about 450 bps and 130 bps were found.
CC   The 450 bp band corresponded to the expected length between the two
CC   primer sites, but the 130 bp band could not be predicted from any
CC   of the previously identified cadherin sequences. The 450 bp and 130
CC   bp bands were extracted and sequenced. Nineteen novel partial cDNA
CC   clones were isolated. The DNA and deduced AA sequences of the
CC   clones (including sequences corresp. to the PCR primers) are given
CC   in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the
CC   cDNA clones are homologous to, but distinct from the known
CC   cadherins. The cadherins described thus far have highly conserved
CC   short AA sequences in the EC-3 including the consensus sequence
CC   D-Y-E or D-E-F located at the middle region of the subdomain and
CC   the consensus sequence in R58879 or R58880 at its end, while the
CC   corresp. sequences of other subdomains, except for the 5th extra-
CC   cellular subdomain (EC-5), are D-R-E and the sequence in R58881
CC   respectively. In contrast the deduced AA sequences of the new
CC   clones that corresp. to cadherin extracellular subdomains include
CC   the sequence D-Y-E or D-F-E at one end, but have the sequence
CC   D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.
CC   The polypeptides encoded by the partial clones are homologous to
CC   previously identified cadherins but did not show significant
CC   homology to any other sequences in Genbank. Therefore, the partial
CC   cDNAs appear to comprise a new subclass of cadherin-related
CC   molecules.
SQ   Sequence   43 AA;

  Query Match             56.8%;  Score 50;  DB 1;  Length 43;
  Best Local Similarity   66.7%;  Pred. No. 0.085;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |:  ||| ||||| |
Db      29 CKIIIKVVDVNDNAP 43


RESULT   8
R87107
ID   R87107 standard; Peptide; 43 AA.
AC   R87107;
DT   28-AUG-1996  (first entry)
DE   Protocadherin clone RAT-224.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW   catenin; therapy; clone.
OS   Rattus rattus.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; U08071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   N-PSDB; T03582.
```

```
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities
PS   Example 1; Page 40; 146pp; English.
CC   R87102-R87120 represent partial fragments of the rat protocadherin
CC   sequence.  The cDNAs encoding these sequences were isolated after
CC   screening a rat brain cDNA preparation with the primers shown in T03575
CC   and T03576.  The primers were constructed from portions of the amino acid
CC   sequences of the third and fourth extracellular domains of published
CC   cadherin sequences.  The full length cDNA sequence encoding rat
CC   protocadherin pc5 is represented in T03574.  The cytoplasmic domain of
CC   cadherin interacts with the cytoskeleton through catenins and other
CC   cytoskeleton associated proteins.  The cytoplasmic domain is not present
CC   in all cadherins, but in those which possess it, it is essential for the
CC   cadherins adhesive function.  The cadherins which do not possess a
CC   cytoplasmic domain appear to function via a different method from those
CC   with a cytoplasmic domain.  These protein sequences are involved in
CC   cell-cell adhesion.  These sequences may have regulatory functions in the
CC   cell, as well as the cell-cell adhesive properties.  Antibodies produced
CC   against these sequences are useful for modulating the binding activity of
CC   these protocadherins, and can be used therapeutically.
SQ   Sequence   43 AA;

  Query Match             56.8%;  Score 50;  DB 1;  Length 43;
  Best Local Similarity   66.7%;  Pred. No. 0.085;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |:  ||| ||||| |
Db      29 CKIIIKVVDVNDNAP 43


RESULT   9
W61273
ID   W61273 standard; Protein; 1180 AA.
AC   W61273;
DT   27-OCT-1998  (first entry)
DE   Homo sapiens protocadherin-4.
KW   pcdh-4 gene; VE-cadherin; protocadherin-4; VE-cadherin-2;
KW   vascular endothelial; angiogenesis; modulation; glaucoma; psoriasis;
KW   inflammatory disease; organ transplantation; treatment; inhibition;
KW   tumour; metastasis; rheumatoid arthritis; diagnosis; detection;
KW   cell adhesion; atherosclerosis; myocardial ischaemia.
OS   Homo sapiens.
PN   WO9825946-A1.
PD   18-JUN-1998.
PF   05-NOV-1997; U20006.
PR   12-DEC-1996; SE-004731.
PA   (IMCL-) IMCLONE SYSTEMS INC.
PI   Dejana E, Telo P;
DR   WPI; 98-348441/30.
DR   N-PSDB; V27593.
PT   New isolated vascular endothelial cadherin-2 - used to develop
PT   products for modulating angiogenesis, e.g. for treating tumours,
PT   glaucoma, psoriasis, inflammatory diseases or organ transplantation
PS   Claim 1; Page 31-37; 54pp; English.
CC   The sequence is that of protocadherin-4 (pcdh-4), or as it may
CC   alternatively be known, VE-cadherin-2. As a vascular endothelial
CC   cadherin it promotes cell-to-cell homotypic adhesion and its
CC   expression is upregulated in proliferating endothelial cells in
CC   comparison to resting cells. It can be used for inhibiting
CC   angiogenesis and inhibiting pathological conditions such as tumours,
CC   neovascular glaucoma, proliferative retinopathy including proliferative
CC   diabetic retinopathy, macular degeneration, hemangiomas, angiofibromas,
CC   and psoriasis. It may also be used for the prevention or inhibition of
CC   leukocyte infiltration, tumour cell metastasis, or endothelial
CC   permeability, as a vaccine and for making endothelial junctions more
CC   permeable to antigens, thus indicating use of the modifiers for
CC   treatment or prevention of acute and chronic inflammatory diseases,
CC   organ transplantation, myocardial ischaemia, atherosclerosis, rheumatoid
CC   arthritis and intestinal infection. It may also be used for detection,
CC   diagnosis and drug screening.
SQ   Sequence   1180 AA;

  Query Match             56.8%;  Score 50;  DB 1;  Length 1180;
  Best Local Similarity   66.7%;  Pred. No. 3.3;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |:  ||| ||||| |
Db     336 CKVLIKVLDVNDNAP 350


RESULT   10
R58875
ID   R58875 standard; Protein; 41 AA.
AC   R58875;
DT   17-APR-1995  (first entry)
DE   Rat-352 cadherin partial sequence.
KW   Cadherin; cell adhesion molecule.
OS   Rattus rattus.
PN   WO9414960-A.
PD   07-JUL-1994.
PF   23-DEC-1993; U12588.
PR   29-DEC-1992; US-998003.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 94-293849/36.
DR   N-PSDB; Q68967.
PT   Polynucleotide sequences encoding new proto:cadherins - useful
PT   for modulating natural binding and regulating activities.
PS   Example; Page 44; 114pp; English.
CC   Two regions of conserved AA sequence, one from the middle of the
CC   third cadherin extracellular subdomain (EC-3) and the other from the
CC   C-terminus of the fourth extracellular subdomain (EC-4) were
CC   identified. The corresp. degenerate oligos (Q68949, Q68950) were
CC   designed for use as PCR primers. PCR was carried out on a rat brain
CC   cDNA prepn. Two major bands of about 450 bps and 130 bps were found.
CC   The 450 bp band corresponded to the expected length between the two
CC   primer sites, but the 130 bp band could not be predicted from any
CC   of the previously identified cadherin sequences. The 450 bp and 130
CC   bp bands were extracted and sequenced. Nineteen novel partial cDNA
CC   clones were isolated. The DNA and deduced AA sequences of the
CC   clones (including sequences corresp. to the PCR primers) are given
CC   in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the
CC   cDNA clones are homologous to, but distinct from the known
CC   cadherins. The cadherins described thus far have highly conserved
CC   short AA sequences in the EC-3 including the consensus sequence
CC   D-Y-E or D-E-F located at the middle region of the subdomain and
CC   the consensus sequence in R58879 or R58880 at its end, while the
CC   corresp. sequences of other subdomains, except for the 5th extra-
CC   cellular subdomain (EC-5), are D-R-E and the sequence in R58881
CC   respectively. In contrast the deduced AA sequences of the new
CC   clones that corresp. to cadherin extracellular subdomains include
CC   the sequence D-Y-E or D-F-E at one end, but have the sequence
CC   D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.
CC   The polypeptides encoded by the partial clones are homologous to
CC   previously identified cadherins but did not show significant
CC   homology to any other sequences in Genbank. Therefore, the partial
CC   cDNAs appear to comprise a new subclass of cadherin-related
CC   molecules.
SQ   Sequence   41 AA;

  Query Match             54.5%;  Score 48;  DB 1;  Length 41;
  Best Local Similarity   60.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |  :|:| ||||| |
Db      27 CTVSIQVVDVNDNAP 41


RESULT   11
R87117
```

```
ID   R87117 standard; Peptide; 41 AA.
AC   R87117;
DT   28-AUG-1996  (first entry)
DE   Protocadherin clone RAT-352.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW   catenin; therapy; clone.
OS   Rattus rattus.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; U08071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   N-PSDB; T03592.
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities.
PS   Example 1; Page 47; 146pp; English.
CC   R87102-R87120 represent partial fragments of the rat protocadherin
CC   sequence.  The cDNAs encoding these sequences were isolated after
CC   screening a rat brain cDNA preparation with the primers shown in T03575
CC   and T03576.  The primers were constructed from portions of the amino acid
CC   sequences of the third and fourth extracellular domains of published
CC   cadherin sequences.  The full length cDNA sequence encoding rat
CC   protocadherin pc5 is represented in T03574.  The cytoplasmic domain of
CC   cadherin interacts with the cytoskeleton through catenins and other
CC   cytoskeleton associated proteins.  The cytoplasmic domain is not present
CC   in all cadherins, but in those which possess it, it is essential for the
CC   cadherins adhesive function.  The cadherins which do not possess a
CC   cytoplasmic domain appear to function via a different method from those
CC   with a cytoplasmic domain.  These protein sequences are involved in
CC   cell-cell adhesion.  These sequences may have regulatory functions in the
CC   cell, as well as the cell-cell adhesive properties.  Antibodies produced
CC   against these sequences are useful for modulating the binding activity of
CC   these protocadherins, and can be used therapeutically.
SQ   Sequence   41 AA;

  Query Match             54.5%;  Score 48;  DB 1;  Length 41;
  Best Local Similarity   60.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |  :|:| ||||| |
Db     27 CTVSIQVVDVNDNAP 41


RESULT  12
W13130
ID   W13130 standard; Protein; 780 AA.
AC   W13130;
DT   13-MAY-1997  (first entry)
DE   Full length human cadherin-5.
KW   Ca2+ dependent; cell adhesion protein; placental; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion.
OS   Homo sapiens.
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   N-PSDB; T61921.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Claim 5; Columns 75-78; 59pp; English.
CC   The present sequence is full length human cadherin-5, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a placental cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   780 AA;

  Query Match             54.5%;  Score 48;  DB 1;  Length 780;
  Best Local Similarity   81.8%;  Pred. No. 4.4;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 IKVKDVNDNFP 15
          ||| ||||| :|
Db    135 IKVHDVNDNWP 145


RESULT  13
W25634
ID   W25634 standard; Protein; 780 AA.
AC   W25634;
DT   03-NOV-1997  (first entry)
DE   Human cadherin-5.
KW   Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW   superfamily; cytoskeleton; eatenin; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 622
FT                   /note= "encoded by AAG"
PN   US5646250-A.
PD   08-JUL-1997.
PF   17-APR-1992; 872643.
PR   19-APR-1993; US-049460.
PR   17-APR-1992; US-872643.
PR   01-NOV-1994; US-332638.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-362997/33.
DR   N-PSDB; T85401.
PT   Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PS   Claim 1; Column 69-72; 56pp; English.
CC   This sequence represents human cadherin-5.  The invention specifically
CC   provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC   cadherin-8, -5, -11 and -13.  Cadherins are calcium-dependent cell
CC   adhesion proteins.  They are glycosylated integral membrane proteins
CC   that have an N-terminal extracellular domain that determines binding
CC   specificity, a hydrophobic membrane spanning region and a C-terminal
CC   cytoplasmic domain, which is highly conserved among members of the
CC   superfamily.  The C-terminal domain interacts with the cytoskeleton
CC   through eatenins and other cytoskeleton-associated proteins.  The
CC   novel cadherin proteins may be used in the analysis of the role of
CC   cadherins in various cancers.  Sequence analysis of the cadherin
CC   proteins also allows investigation of the structure and function of
CC   cadherin.  The cadherin proteins may be isolated by using anti-cadherin
CC   antibodies.  These antibodies may also be used to modulate the activity
CC   of cadherin and to determine the tissue specific distribution of cadherin
CC   proteins.  Each subclass of cadherins has a unique tissue distribution
CC   pattern.
SQ   Sequence   780 AA;

  Query Match             54.5%;  Score 48;  DB 1;  Length 780;
  Best Local Similarity   81.8%;  Pred. No. 4.4;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 IKVKDVNDNFP 15
          ||| ||||| :|
Db    135 IKVHDVNDNWP 145


RESULT  14
W13126
ID   W13126 standard; Protein; 799 AA.
AC   W13126;
```

```
DT   13-MAY-1997  (first entry)
DE   Full length rat cadherin-8.
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion.
OS   Rattus rattus.
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   N-PSDB; T61917.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Example 1; Columns 49-52; 59pp; English.
CC   The present sequence is full length rat cadherin-8, which is a
CC   Ca2+ dependent cell adhesion protein. The rat cadherin cDNA was
CC   isolated from a brain cDNA library using a labelled rat cadherin-8
CC   extracellular domain PCR fragment as a probe. The rat cDNA was then
CC   used to isolate the cDNA encoding its full length human homologue
CC   from a human foetal brain cDNA library.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   799 AA;


  Query Match             52.3%;  Score 46;  DB 1;  Length 799;
  Best Local Similarity   72.7%;  Pred. No. 9.7;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     155 IKVQDINDNAP 165


RESULT  15
W13132
ID   W13132 standard; Protein; 793 AA.
AC   W13132;
DT   14-MAY-1997  (first entry)
DE   Full length human cadherin-8.
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion.
OS   Homo sapiens.
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   N-PSDB; T61923.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Example 2; Columns 85-90; 59pp; English.
CC   The present sequence is full length human cadherin-8, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a foetal brain cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   793 AA;


  Query Match             52.3%;  Score 46;  DB 1;  Length 793;
  Best Local Similarity   72.7%;  Pred. No. 9.6;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     148 IKVQDINDNAP 158


Search completed: January  6, 2000, 15:40:37
Job time: 1684 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:37 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-6
Perfect score: 82
Sequence: 1 SARTLNNRYTGPYTF 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	100.0	999	1	R30742	Human pemphigus vu
2	82	100.0	614	1	W07908	Pemphigus vulgaris
3	82	100.0	15	1	W04846	Self epitope of de
4	38	46.3	431	1	R97293	Mouse CRF RB1 rece
5	37	45.1	431	1	R90575	Rat CRF2-beta rece
6	37	45.1	454	1	R93616	Kaposi's sarcoma a
7	37	45.1	454	1	R97842	Kaposi's sarcoma a
8	36.5	44.5	820	1	R71802	N-acetylheparosan
9	36	43.9	516	1	R12098	Haem 84-1 portion
10	36	43.9	480	1	R12099	Haem 84-2 portion
11	36	43.9	3898	1	R10473	Hog cholera virus
12	36	43.9	1418	1	W15264	Bloom syndrome act
13	36	43.9	185	1	W31545	Bloom's syndrome B
14	36	43.9	271	1	W31546	Bloom's syndrome B
15	36	43.9	515	1	W31547	Bloom's syndrome B
16	36	43.9	1417	1	W31548	Bloom's syndrome B
17	36	43.9	739	1	W31549	Bloom's syndrome B
18	36	43.9	1417	1	W31550	Bloom's syndrome B
19	36	43.9	1417	1	W31551	Bloom's syndrome B
20	36	43.9	244	1	W77619	Mercuric reductase
21	36	43.9	344	1	W98545	H. pylori GHPO 156
22	36	43.9	1261	1	W93601	Tomato Xa21 clone
23	35	42.7	325	1	R56522	Protein kinase (CK
24	35	42.7	364	1	R56523	Protein kinase (CK
25	35	42.7	337	1	R56524	Protein kinase (CK
26	35	42.7	369	1	R62443	R.CviJI encoded by
27	35	42.7	364	1	R76619	Human HRR25-like c
28	35	42.7	337	1	R76620	Human HRR25-like c
29	35	42.7	325	1	R76618	Human HRR25-like c
30	35	42.7	1196	1	R75189	Osteoinductive ret
31	35	42.7	693	1	R88646	Neisseria meningit
32	35	42.7	610	1	R88654	N.meningitidis 694
33	35	42.7	752	1	R97199	Chimaeric thermost
34	35	42.7	429	1	W08606	Chimeric MoMLV and
35	35	42.7	449	1	W08608	Chimeric MoMLV and
36	35	42.7	428	1	W08609	Chimeric MoMLV and
37	35	42.7	105	1	W52243	Antibody LD1/2-6-3
38	35	42.7	105	1	W52241	Antibody LD1/2-6-3
39	35	42.7	933	1	W98599	H. pylori GHPO 127
40	35	42.7	623	1	W30620	Arabidopsis La-er
41	35	42.7	623	1	W30622	Arabidopsis enhanc
42	35	42.7	623	1	W30623	Arabidopsis enhanc
43	35	42.7	623	1	W30624	Arabidopsis mutant
44	35	42.7	1079	1	Y03163	MuLV reverse trans
45	34	41.5	354	1	R42424	Rat gustducin alph

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 82; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SARTLNNRYTGPYTF 15
||||||||||||||||
Db 512 SARTLNNRYTGPYTF 526

RESULT 2
W07908
ID W07908 standard; protein; 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphis vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease-causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucuous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose

```
CC   pemphigus vulgaris.
SQ   Sequence   614 AA;


  Query Match             100.0%;  Score 82;  DB 1;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SARTLNNRYTGPYTF 15
          |||||||||||||||
Db    511 SARTLNNRYTGPYTF 525


RESULT   3
W04846
ID   W04846 standard; peptide; 15 AA.
AC   W04846;
DT   18-FEB-1997  (first entry)
DE   Self epitope of desmoglein 3, implicated in autoimmune disease.
KW   Tolerisation; self-epitope; antigen; autoimmune disease;
KW   autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW   pemphigus vulgaris; desmoglein; multiple sclerosis;
KW   herpes simplex virus; adenovirus; phosphomannomutase;
KW   human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW   influenza; haemagglutinin; reovirus; sigma protein.
OS   Homo sapiens.
PN   WO9627387-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03182.
PR   07-MAR-1995; US-400796.
PA   (HARD ) HARVARD COLLEGE.
PI   Strominger JL,  Wucherpfennig KW;
DR   WPI; 96-425218/42.
PT   Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT   antigens - useful in disease treatment, and method for
PT   identification of other self and non-self antigens implicated in
PT   auto-immune disease
PS   Claim 1; Page 41; 58pp; English.
CC   Pharmaceutical preparations for tolerisation to antigens comprise
CC   either an isolated human non-collagen or non-myslin basic protein
CC   (MBP) polypeptide which is capable of tolerising an individual to an
CC   autoantigen; or an isolated human pathogen polypeptide capable of
CC   tolerising an individual to that polypeptide.  In both cases, the
CC   polypeptide (whether self or non-self) includes an amino acid
CC   sequence corresponding to a sequence motif for a MHC class II
CC   protein, such as HLA-DR, which is associated with a human autoimmune
CC   disease and which binds to the polypeptide to activate autoreactive
CC   T-cells in individuals with the autoimmune disease. This peptide is
CC   derived from the human desmoglein 3 protein (amino acids 512-526)
CC   and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC   derived from the human desmoglein protein are described in W04841-47.
SQ   Sequence   15 AA;


  Query Match             100.0%;  Score 82;  DB 1;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-08;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SARTLNNRYTGPYTF 15
          |||||||||||||||
Db      1 SARTLNNRYTGPYTF 15


RESULT   4
R97293
ID   R97293 standard; Protein; 431 AA.
AC   R97293;
DT   21-AUG-1996  (first entry)
DE   Mouse CRF RB1 receptor.
KW   Corticotropin releasing factor receptor; CRF-R; corticoliberin;
KW   signal transduction.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   modified_site   52
FT                   /label= N-glycosylation_site
FT   modified_site   61
FT                   /label= N-glycosylation_site
FT   modified_site   94
FT                   /label= N-glycosylation_site
FT   modified_site   105
FT                   /label= N-glycosylation_site
FT   modified_site   113
FT                   /label= N-glycosylation_site
FT   domain          139. .159
FT                   /label= Transmembrane_domain-1
FT   domain          169. .188
FT                   /label= Transmembrane_domain-2
FT   domain          206. .229
FT                   /label= Transmembrane_domain-3
FT   domain          245. .265
FT                   /label= Transmembrane_domain-4
FT   domain          285. .307
FT                   /label= Transmembrane_domain-5
FT   domain          331. .351
FT                   /label= Transmembrane_domain-6
FT   domain          366. .385
FT                   /label= Transmembrane_domain-7
PN   WO9617934-A2.
PD   13-JUN-1996.
PF   06-DEC-1995; U15909.
PR   09-DEC-1994; US-353537.
PR   17-JAN-1995; US-374009.
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
PI   Chen R,  Donaldson CJ,  Lewis KA,  Perrin MH,  Sawchenko P;
PI   Vale WW;
DR   WPI; 96-287179/29.
DR   N-PSDB; T28972.
PT   Isolated corticotropin-releasing factor receptor (CRF-R) - used to
PT   develop prods. for modulating signal transduction activity mediated
PT   by CRF-R
PS   Claim 3; Page 83-85; 102pp; English.
CC   Mouse corticotropin releasing factor receptor mCRF-RB1 was
CC   identified as the product of a cDNA clone (T28972) isolated from a
CC   mouse heart library.  Recombinant mCRF-RB1 can be expressed in
CC   host cells transformed by the cDNA clone.  The receptor can be used
CC   to identify agonists and antagonists that modulate the signal
CC   transduction activity mediated by CRF receptors.  It may be
CC   administered therapeutically to reduce high ACTH levels caused by
CC   excess CRF.
SQ   Sequence   431 AA;


  Query Match             46.3%;  Score 38;  DB 1;  Length 431;
  Best Local Similarity   46.2%;  Pred. No. 49;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 RTLNNRYTGPYTF 15
          ||    ::|||:
Db     47 RTTIGNFSGPYTY 59


RESULT   5
R90575
ID   R90575 standard; Protein; 431 AA.
AC   R90575;
DT   08-APR-1996  (first entry)
DE   Rat CRF2-beta receptor.
KW   CRF2-beta receptor; corticotropin-releasing factor-2 receptor;
KW   cerebrovascular disorder; memory disorder; Alzheimer disease.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   domain          1. .117
FT                   /label= Extracellular_N-terminal_domain
FT   domain          118. .138
FT                   /label= Transmembrane_domain
FT   domain          139. .147
```

```
FT                      /label= Intracellular_domain
FT   domain             148. .167
FT                      /label= Transmembrane_domain
FT   domain             168. .184
FT                      /label= Extracellular_domain
FT   domain             185. .208
FT                      /label= Transmembrane_domain
FT   domain             209. .223
FT                      /label= Intracellular_domain
FT   domain             224. .244
FT                      /label= Transmembrane_domain
FT   domain             245. .261
FT                      /label= Extracellular_domain
FT   domain             262. .286
FT                      /label= Transmembrane_domain
FT   domain             287. .309
FT                      /label= Intracellular_domain
FT   domain             310. .329
FT                      /label= Transmembrane_domain
FT   domain             330. .342
FT                      /label= Extracellular_domain
FT   domain             343. .363
FT                      /label= Transmembrane_domain
FT   domain             364. .431
FT                      /label= C-terminal_intracellular_domain
PN   WO9534651-A2.
PD   21-DEC-1995.
PF   14-JUN-1995; U07757.
PR   14-JUN-1994; US-259959.
PR   31-JAN-1995; US-381433.
PR   07-JUN-1995; US-485984.
PA   (NEUR-) NEUROCRINE BIOSCIENCES INC.
PI   Chalmers D,  De Souza EB,  Grigoriadis DE,  Liaw CW;
PI   Lovenberg TW, Oltersdorf T;
DR   WPI; 96-049680/05.
DR   N-PSDB; T12244.
PT   Corticotropin-releasing factor-2 receptor, and DNA encoding it -
PT   used to isolate CRF-2 receptor antagonists for the treatment of
PT   cerebrovascular disorders, memory disorders and Alzheimer's disease
PS   Disclosure; Page 63-66; 109pp; English.
CC   Rat corticotropin-releasing factor-2-beta (CRF2-beta) receptor
CC   (R90575) is a membrane-bound G-coupled protein receptor involved
CC   in signal transduction.  It can be produced by expression of
CC   encoding cDNA (T12244) in procaryotic or eucaryotic host cells.
CC   Recombinant CRF2 receptor is used to screen CRF2 receptor agonists
CC   and antagonists of therapeutic appln., and to prepare antibodies
CC   which specifically bind to CRF2 receptors.
SQ   Sequence   431 AA;


  Query Match             45.1%;  Score 37;  DB 1;  Length 431;
  Best Local Similarity   38.5%;  Pred. No. 73;
  Matches    5;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 RTLNNRYTGPYTF 15
           ||    ::|||::
Db      47 RTTTRNFSGPYSY 59


RESULT   6
R93616
ID   R93616 standard; Protein; 454 AA.
AC   R93616;
DT   13-AUG-1996  (first entry)
DE   Kaposi's sarcoma associated herpesvirus ORF32 product.
KW   Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis;
KW   therapy.
OS   Kaposi's sarcoma associated herpesvirus.
PN   WO9606159-A1.
PD   29-FEB-1996.
PF   11-AUG-1995; U10194.
PR   18-AUG-1994; US-292365.
PR   21-NOV-1994; US-343100.
PR   11-APR-1995; US-420235.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PA   (GRAN/) GRANT D E.
PA   (VIEL/) VIELE L.
PI   Chang Y,  Moore PS,  Grant DE,  Viele L;
DR   WPI; 96-151362/15.
DR   N-PSDB; T16822.
PT   Herpes:virus DNA associated with Kaposi's sarcoma - also associated
PT   vectors and proteins, used in detection and vaccination.
PS   Claim 17; Page 235-237; 305pp; English.
CC   Kaposi's sarcoma associated herpes virus (KSHV) clone KS5 (T16806),
CC   obtd. from a KS lesion genomic library, includes 15 complete ORFs and
CC   2 incomplete ORFs (T16807-23) named according to their herpesvirus
CC   saimiri positional homologues.  The ORF32 (T16822) product is
CC   given in R93616.  KSHV proteins and peptides may be obtd. by
CC   incorporating encoding sequences into a vector and expression in
CC   host cells.  They are useful in vaccines or for raising antibodies of
CC   diagnostic or therapeutic value.
SQ   Sequence   454 AA;


  Query Match             45.1%;  Score 37;  DB 1;  Length 454;
  Best Local Similarity   54.5%;  Pred. No. 77;
  Matches    6;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ARTLNNRYTGP 12
           |  :| || ||
Db       3 AHAINERYVGP 13


RESULT   7
R97842
ID   R97842 standard; Protein; 454 AA.
AC   R97842;
DT   11-SEP-1996  (first entry)
DE   Kaposi's sarcoma associated herpesvirus ORF32 product.
KW   Kaposi's sarcoma; gamma-2 herpesvirus; KSHV; therapy; diagnosis;
KW   vaccine; diagnosis; AIDS.
OS   Kaposi's sarcoma associated herpesvirus.
PN   WO9615779-A1.
PD   30-MAY-1996.
PF   21-NOV-1995; U15138.
PR   21-NOV-1994; US-343101.
PR   11-APR-1995; US-420235.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PI   Chang Y,  Moore PS;
DR   WPI; 96-268320/27.
DR   N-PSDB; T30694.
PT   Herpes virus associated with Kaposi's sarcoma - also definitive DNA
PT   sequences, useful for diagnosis of and to develop prods. for
PT   treatment of Kaposi's sarcoma
PS   Claim 17; Page 188-189; 277pp; English.
CC   Lambda clone KS5 (T30681) is a fragment of a newly identified human
CC   gamma-2 herpesvirus associated with Kaposi's sarcoma (KS).  KS5 has
CC   17 open reading frames (T30682-98), 15 of which are complete,
CC   including ORF32 (T30694).  The protein products (R97830-46,
CC   respectively) of the 17 ORFs can be expressed in eukaryotic or
CC   bacterial host cells for use as vaccines, for KS diagnosis, or for
CC   raising antibodies.
SQ   Sequence   454 AA;


  Query Match             45.1%;  Score 37;  DB 1;  Length 454;
  Best Local Similarity   54.5%;  Pred. No. 77;
  Matches    6;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ARTLNNRYTGP 12
           |  :| || ||
Db       3 AHAINERYVGP 13


RESULT   8
R71802
```

```
ID   R71802 standard; Protein; 820 AA.
AC   R71802;
DT   19-OCT-1995  (first entry)
DE   N-acetylheparosan fragmentation enzyme.
KW   N-acetylheparosan; polysaccharide K5; fragmentation; degradation;
KW   pharmaceutical starting material.
OS   Escherichia coli (K5) SEBR 3282 (CNCM I-1013).
FH   Key             Location/Qualifiers
FT   binding_site    384. .387
FT                   /note= "Asn-X1-X2-Ser motif (X1=Ile, Leu or Tyr
FT                   and X2=Val, Ile or Ala)"
FT   binding_site    411. .414
FT                   /note= "Asn-X1-X2-Ser motif (X1=Ile, Leu or Tyr
FT                   and X2=Val, Ile or Ala)"
FT   binding_site    433. .436
FT                   /note= "Asn-X1-X2-Ser motif (X1=Ile, Leu or Tyr
FT                   and X2=Val, Ile or Ala)"
FT   binding_site    461. .464
FT                   /label= potential_lipoprotein_attachment_site
FT                   /note= "Asn-X1-X2-Ser motif (X1=Ile, Leu or Tyr
FT                   and X2=Val, Ile or Ala)"
FT   binding_site    495. .498
FT                   /note= "Asn-X1-X2-Ser motif (X1=Ile, Leu or Tyr
FT                   and X2=Val, Ile or Ala)"
FT   region          804. .809
FT                   /note= "strongly charged C-terminal region"
PN   FR2709132-A.
PD   24-FEB-1995.
PF   17-AUG-1993; 010050.
PR   17-AUG-1993; FR-010050.
PA   (SNFI ) ELF SANOFI.
PI   Legoux R, Lelong P, Salome MLV;
DR   WPI; 95-100554/14.
DR   N-PSDB; Q86268.
PT   New DNA encoding enzyme for fragmenting N-acetyl:heparosan -
PT   giving products useful as pharmaceutical starting material, also
PT   related protein, vector, transformed cells, etc.
PS   Claim 2; Page 38-40; 54pp; French.
CC   An enzyme which degrades high mol.wt. N-acetylheparosan into
CC   fragments of lower mol.wt. (more suitable for use as pharmaceutical
CC   starting materials) can be obtained from E.coli (K5) SEBR 3282. The
CC   DNA sequence (Q86268) coding for the enzyme was isolated in a
CC   plasmid (p838.7) which has been deposited in E.coli RR1 strain as
CC   CNCM I-1352. The deduced amino acid sequence (R71802) has homology
CC   with an exo-poly-alpha-D-galacturonidasse from Erwinia chrysanthemi.
SQ   Sequence   820 AA;

  Query Match             44.5%;  Score 36.5;  DB 1;  Length 820;
  Best Local Similarity   47.1%;  Pred. No. 1.8e+02;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    7;  Gaps    1;

Qy       6 NNRYTG-------PYTF 15
           :| |||       ||||
Db       9 HNEYTGNGVTTSFPYTF 25


RESULT   9
R12098
ID   R12098 standard; Protein; 516 AA.
AC   R12098;
DT   22-JUL-1991  (first entry)
DE   Haem 84-1 portion of H.contortus 55A surface protein.
KW   Parasitic nematode; vaccine; Hc.
OS   Haemonchus contortus.
PN   AU9062569-A.
PD   21-MAR-1991.
PF   17-SEP-1990; 062569.
PR   18-SEP-1989; US-408339.
PR   01-MAR-1990; US-487181.
PA   (SYNE-) SYNERGEN INC.
DR   WPI; 91-133285/19.
DR   N-PSDB; Q11798.
PT   Protein from Haemonchus contortus and other nematodes - used as
PT   therapeutic and prophylactic agent to protect plants, animals or
PT   humans from parasitic nematode infection.
PS   Disclosure; Fig 49; 209pp; English.
CC   The proteins derived from the nematode DNA may be used to derive
CC   vaccines against parasitic infection of plants, humans and animals
CC   esp. sheep. MAbs may also be raised to provide passive therapy and
CC   prophylaxis against infection.
SQ   Sequence   516 AA;

  Query Match             43.9%;  Score 36;  DB 1;  Length 516;
  Best Local Similarity   54.5%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 LNNRYTGPYTF 15
           :|||| | | :
Db     484 VNNRYGGRYKY 494


RESULT  10
R12099
ID   R12099 standard; Protein; 480 AA.
AC   R12099;
DT   22-JUL-1991  (first entry)
DE   Haem 84-2 portion of H.contortus 55A surface protein.
KW   Parasitic nematode; vaccine; Hc.
OS   Haemonchus contortus.
PN   AU9062569-A.
PD   21-MAR-1991.
PF   17-SEP-1990; 062569.
PR   18-SEP-1989; US-408339.
PR   01-MAR-1990; US-487181.
PA   (SYNE-) SYNERGEN INC.
DR   WPI; 91-133285/19.
DR   N-PSDB; Q11799.
PT   Protein from Haemonchus contortus and other nematodes - used as
PT   therapeutic and prophylactic agent to protect plants, animals or
PT   humans from parasitic nematode infection.
PS   Disclosure; Fig 49; 209pp; English.
CC   The proteins derived from the nematode DNA may be used to derive
CC   vaccines against parasitic infection of plants, humans and animals
CC   esp. sheep. MAbs may also be raised to provide passive therapy and
CC   prophylaxis against infection.
SQ   Sequence   480 AA;

  Query Match             43.9%;  Score 36;  DB 1;  Length 480;
  Best Local Similarity   54.5%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 LNNRYTGPYTF 15
           :|||| | | :
Db     376 VNNRYGGRYKY 386


RESULT  11
R10473
ID   R10473 standard; Protein; 3898 AA.
AC   R10473;
DT   09-APR-1991  (first entry)
DE   Hog cholera virus genome product.
KW   Swine fever virus; HCV; pestvirus; border disease virus; BDV;
KW   Bovine viral diarrhoea virus; BVDV; Togaviridae; ss.
OS   Hog Cholera Virus.
PN   WO9100352-A.
PD   10-JAN-1991.
PF   29-JUN-1990; NL0092.
PR   29-JUN-1989; NL-001651.
PA   (DIER-) CENT DIERGENEESKUND.
PI   Moorman R, Wensvoort G;
DR   WPI; 91-036746/05.
DR   P-PSDB; Q10353.
```

```
PT   Protection of animals against microbial infection - using
PT   nucleotide sequences related to the microbe and a stop transfer
PT   region
PS   Claim 15; Fig 1; 33pp; English.
CC   The genome sequence and its encoded product are sufficiently similar
CC   to other pestviruses of the family Togaviridae, that they may be
CC   used in vaccination and diagnosis of pestvirus infection.
SQ   Sequence   3898 AA;


  Query Match             43.9%;  Score 36;  DB 1;  Length 3898;
  Best Local Similarity   63.6%;  Pred. No. 1.3e+03;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ARTLNNRYTGP 12
           |:|| |||  |
Db     988 AKTLRNRYYEP 998


RESULT  12
W15264
ID   W15264 standard; Protein; 1418 AA.
AC   W15264;
DT   26-JAN-1998  (first entry)
DE   Bloom syndrome active BLM protein.
KW   BLM; Bloom syndrome; BS; mutant; probe; PCR primer; cancer;
KW   therapy; diagnosis; vector.
OS   Homo sapiens.
PN   WO9717979-A1.
PD   22-MAY-1997.
PF   15-NOV-1996; U19046.
PR   15-NOV-1995; US-559303.
PA   (NYBL-) NEW YORK BLOOD CENT INC.
PI   Ellis N,  German J,  Groden J;
DR   WPI; 97-289051/26.
DR   N-PSDB; T67013.
PT   Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT   genes - for gene therapy with nucleic acid encoding active BLM
PT   protein to treat Bloom's syndrome and cancer in general
PS   Claim 58; Fig 2; 51pp; English.
CC   This active BLM protein is encoded by a 4437 bp BLM gene sequence.
CC   This is used in diagnosing Bloom's syndrome where 2 mutated BLM genes
CC   or the absence of a wild-type BLM gene in the nucleic acid of a subject
CC   is detected. The BLM gene is in the 250 kb region between two markers
CC   binding a 1.3 cM region on chromosome 15. cDNA from this region was
CC   isolated, a 847 bp fragment selected and used to screen a HeLa cDNA
CC   library. The longest clone H1 was isolated and extended by PCR to isolate
CC   this BLM gene sequence. Delivery of a functional BLM gene to bone marrow
CC   cells, e.g. via the viral vectors, is used to treat or prevent the onset
CC   of Bloom's syndrome. Identification of the BLM gene and its products
CC   should assist in the development of therapeutic and diagnostic agents for
CC   cancer.
SQ   Sequence   1418 AA;


  Query Match             43.9%;  Score 36;  DB 1;  Length 1418;
  Best Local Similarity   87.5%;  Pred. No. 4.2e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SARTLNNR 8
           |||||||:
Db      17 SARTLNNK 24


RESULT  13
W31545
ID   W31545 standard; Protein; 185 AA.
AC   W31545;
DT   27-JAN-1998  (first entry)
DE   Bloom's syndrome BLM mutated protein.
KW   BLM; Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Protein         1. .185
FT                   /note= "Truncated protein. The deletion of CAA at
FT                          nucleotide positions 631-633 of the wild
FT                          type BLM gene results in a stop codon at
FT                          amino acid position 186"
PN   WO9717979-A1.
PD   22-MAY-1997.
PF   15-NOV-1996; U19046.
PR   15-NOV-1995; US-559303.
PA   (NYBL-) NEW YORK BLOOD CENT INC.
PI   Ellis N,  German J,  Groden J;
DR   WPI; 97-289051/26.
DR   N-PSDB; T93389.
PT   Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT   genes - for gene therapy with nucleic acid encoding active BLM
PT   protein to treat Bloom's syndrome and cancer in general
PS   Disclosure; Page -; 51pp; English.
CC   This is a truncated BLM protein encoded by a mutant gene sequence
CC   isolated from a Japanese Bloom's Syndrome sufferer designated
CC   "97(AsOk)". The deletion of CAA from positions 631-633 of the
CC   wild-type gene results in a stop codon at amino acid position 186.
CC   This was one of the seven unique mutations which were identified in
CC   a study of 10 people with Bloom's Syndrome. Based on the various
CC   mutations, diagnostic tests for Bloom's Syndrome have been developed
CC   which use standard sequence analysis techniques to detect the presence
CC   of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC   Note: The present sequence does not appear in the specification; it
CC   has been made by modifying the H1-5' wild type BLM sequence which is
CC   provided in Figure 2 (W15264).
SQ   Sequence   185 AA;


  Query Match             43.9%;  Score 36;  DB 1;  Length 185;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SARTLNNR 8
           |||||||:
Db      17 SARTLNNK 24


RESULT  14
W31546
ID   W31546 standard; Protein; 271 AA.
AC   W31546;
DT   27-JAN-1998  (first entry)
DE   Bloom's syndrome BLM mutated protein.
KW   BLM; Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Protein         1. .271
FT                   /note= "Truncated protein. The substitution of the base A
FT                          at the position 888 of the wild type H1-5' gene
FT                          to the base T results in a stop codon at amino
FT                          acid position 272"
PN   WO9717979-A1.
PD   22-MAY-1997.
PF   15-NOV-1996; U19046.
PR   15-NOV-1995; US-559303.
PA   (NYBL-) NEW YORK BLOOD CENT INC.
PI   Ellis N,  German J,  Groden J;
DR   WPI; 97-289051/26.
DR   N-PSDB; T93390.
PT   Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT   genes - for gene therapy with nucleic acid encoding active BLM
PT   protein to treat Bloom's syndrome and cancer in general
PS   Disclosure; Page -; 51pp; English.
CC   This is a truncated BLM protein encoded by a mutant gene sequence
CC   isolated from a German Bloom's Syndrome sufferer designated "112(NaSch)".
CC   The substitution of the base A at the position 888 of the wild type H1-5'
CC   gene to the base T results in a stop codon at amino acid position 272.
CC   This was one of the seven unique mutations which were identified in
CC   a study of 10 people with Bloom's Syndrome. Based on the various
```

```
CC   mutations, diagnostic tests for Bloom's Syndrome have been developed
CC   which use standard sequence analysis techniques to detect the presence
CC   of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC   Note: The present sequence does not appear in the specification; it
CC   has been made by modifying the H1-5' wild type BLM sequence which is
CC   provided in Figure 2 (W15264).
SQ   Sequence   271 AA;


  Query Match            43.9%;  Score 36;  DB 1;  Length 271;
  Best Local Similarity  87.5%;  Pred. No. 65;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SARTLNNR 8
          |||||||:
Db     17 SARTLNNK 24


RESULT  15
W31547
ID   W31547 standard; Protein; 515 AA.
AC   W31547;
DT   27-JAN-1998  (first entry)
DE   Bloom's syndrome BLM mutated protein.
KW   BLM; Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Protein         1. .515
FT                   /note= "The insertion of a bp A at position 1610 of
FT                          the wild type H1-5' BLM gene results in  the
FT                          insertion of a novel codon for Lys after amino
FT                          acid position 514 followed by a stop codon"
PN   WO9717979-A1.
PD   22-MAY-1997.
PF   15-NOV-1996; U19046.
PR   15-NOV-1995; US-559303.
PA   (NYBL-) NEW YORK BLOOD CENT INC.
PI   Ellis N,  German J,  Groden J;
DR   WPI; 97-289051/26.
DR   N-PSDB; T93391.
PT   Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT   genes - for gene therapy with nucleic acid encoding active BLM
PT   protein to treat Bloom's syndrome and cancer in general
PS   Disclosure; Page -; 51pp; English.
CC   This is a truncated BLM protein encoded by a mutant gene sequence
CC   isolated from a Japanese Bloom's Syndrome sufferer designated
CC   "93(YoYa)". The insertion of a bp A at the position 1610 in the wild type
CC   H1-5' BLM gene results in the insertion of a novel codon for Lysine after
CC   amino acid position 514 followed by a stop codon.
CC   This was one of the seven unique mutations which were identified in
CC   a study of 10 people with Bloom's Syndrome. Based on the various
CC   mutations, diagnostic tests for Bloom's Syndrome have been developed
CC   which use standard sequence analysis techniques to detect the presence
CC   of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC   Note: The present sequence does not appear in the specification; it
CC   has been made by modifying the H1-5' wild type BLM sequence which is
CC   provided in Figure 2 (W15264).
SQ   Sequence   515 AA;


  Query Match            43.9%;  Score 36;  DB 1;  Length 515;
  Best Local Similarity  87.5%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SARTLNNR 8
          |||||||:
Db     17 SARTLNNK 24


Search completed: January  6, 2000, 15:40:38
Job time: 1685 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:44 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-6
Perfect score: 82
Sequence: 1 SARTLNNRYTGPYTF 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	100.0	999	1	IJHUG3	desmoglein 3 precu
2	46	56.1	531	2	C70333	hypothetical prote
3	42	51.2	163	2	B70854	hypothetical prote
4	40	48.8	848	4	A44282	retrovirus-related
5	39	47.6	2261	1	A42548	genome polyprotein
6	39	47.6	2262	2	S16664	large protein L -
7	39	47.6	282	2	S43577	C28A5.4 protein (c
8	39	47.6	973	2	T01862	hypothetical prote
9	39	47.6	431	2	S25196	mitochondrial impo
10	38	46.3	865	2	S69044	hypothetical prote
11	38	46.3	431	2	I49279	sauvagine/corticot
12	38	46.3	356	2	F71151	probable glucose-1
13	38	46.3	244	2	C70912	hypothetical prote
14	38	46.3	423	2	B70931	probable PPE prote
15	38	46.3	332	2	E69312	nitrate reductase
16	38	46.3	430	2	A56726	corticoliberin rec
17	38	46.3	431	2	I49149	CRF receptor - mou
18	38	46.3	352	2	A75098	glucose-1-phosphat
19	37.5	45.7	690	2	T01183	hypothetical prote
20	37	45.1	387	2	A45117	aspartic proteinas
21	37	45.1	1186	2	S70430	hypothetical prote
22	37	45.1	368	2	A69275	3-ketoacyl-CoA thi
23	37	45.1	302	2	F64883	probable transcrip
24	37	45.1	218	2	H69005	hypothetical prote
25	37	45.1	405	2	A69206	hypothetical prote
26	37	45.1	413	2	S28066	escl protein - fis
27	37	45.1	643	2	S46723	arginine--tRNA lig
28	37	45.1	492	2	S54536	probable membrane
29	37	45.1	331	2	E75067	aminopeptidase fro
30	36.5	44.5	980	2	T00045	cellodextrin phosp
31	36	43.9	2269	1	JQ1750	genome polyprotein
32	36	43.9	362	1	VVVP1S	coat protein VP1 -
33	36	43.9	372	1	VVVPK1	coat protein VP1 -
34	36	43.9	3898	1	GNWVHB	genome polyprotein
35	36	43.9	1237	1	GNFF42	retrovirus-related
36	36	43.9	334	1	G8BPT4	baseplate protein
37	36	43.9	334	2	C70673	probable gpdA2 pro
38	36	43.9	644	2	S44478	NADH dehydrogenase
39	36	43.9	554	2	T01240	probable diphenol
40	36	43.9	1417	2	A57570	Bloom's syndrome r
41	36	43.9	344	2	H64564	UDP-glucose 4-epim
42	36	43.9	344	2	G71858	udp-glucose 4-epim
43	36	43.9	362	2	S29968	coat protein VP1 -
44	36	43.9	699	2	T05225	extensin homolog F
45	36	43.9	391	2	T04609	hypothetical prote

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 82; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SARTLNNRYTGPYTF 15
           |||||||||||||||
Db     512 SARTLNNRYTGPYTF 526
```

RESULT 2
C70333
hypothetical protein aq_372 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
C;Accession: C70333
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70333

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-531 <AQF>
A;Cross-references: GB:AE000686; NID:g2983038; PID:g2983043; GB:AE000657
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_372

Query Match 56.1%; Score 46; DB 2; Length 531;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       1 SARTLNNRYTGPYTF 15
           |  || ||   ||||
Db      77 SIETLKNRLENPYTF 91
```

RESULT 3
B70854
hypothetical protein Rv2991 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C;Accession: B70854
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70854
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-163 <COL>
A;Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PID:e1237736; PID:g2791588
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2991

Query Match 51.2%; Score 42; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       8 RYTGPYT 14
           |||||||
Db     112 RYTGPYT 118
```

RESULT 4
A44282
retrovirus-related pol polyprotein pseudogene - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 04-Jan-1996 #text_change 04-Jan-1996
C;Accession: A44282
R;Hirose, Y.; Takamatsu, M.; Harada, F.
Virology 192, 52-61, 1993
A;Title: Presence of env genes in members of the RTVL-H family of human endogenous retro
A;Reference number: A44282; MUID:93297138
A;Accession: A44282
A;Status: conceptual translation of pseudogene
A;Molecule type: mRNA
A;Residues: 1-848 <HIR>
A;Cross-references: GB:D10083
C;Keywords: pseudogene
F;45/Region: opal stop codon
F;485/Region: opal stop codon
F;694/Region: ochre stop codon
F;815/Region: opal stop codon
F;827/Region: ochre stop codon
F;835/Region: opal stop codon

Query Match 48.8%; Score 40; DB 4; Length 848;
Best Local Similarity 63.6%; Pred. No. 42;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy       3 RTLNNRYTGPY 13
           :||  |:||||
Db     660 QTLQPRWTGPY 670
```

RESULT 5
A42548
genome polyprotein - mumps virus (strain Miyahara)
N;Alternate names: L protein
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: mumps virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jun-1999
C;Accession: A42548
R;Okazaki, K.; Tanabayashi, K.; Takeuchi, K.; Hishiyama, M.; Okazaki, K.; Yamada, A.
Virology 188, 926-930, 1992
A;Title: Molecular cloning and sequence analysis of the mumps virus gene encoding the
A;Reference number: A42548; MUID:92263804
A;Accession: A42548
A;Molecule type: genomic RNA
A;Residues: 1-2261 <OKA>
A;Cross-references: GB:D10575; NID:g222145; PIDN:BAA01432.1; PID:d1001906; PID:g22214
C;Genetics:
A;Gene: L
C;Superfamily: parainfluenza virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase

Query Match 47.6%; Score 39; DB 1; Length 2261;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2 ARTLNNRYTGPYTF 15
           |||||: |  |: |
Db     686 ARTLNSMYGIPHLF 699
```

RESULT 6
S16664
large protein L - parainfluenza virus type 2
C;Species: parainfluenza virus type 2
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Jun-1999
C;Accession: S16664
R;Kawano, M.; Okamoto, K.; Bando, H.; Kondo, K.; Tsurudome, M.; Komada, H.; Nishio, M
Nucleic Acids Res. 19, 2739-2746, 1991
A;Title: Characterizations of the human parainfluenza type 2 virus gene encoding the
A;Reference number: S16659; MUID:91252221
A;Accession: S16664
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2262 <KAW>
A;Cross-references: EMBL:X57559; NID:g61985; PIDN:CAA40788.1; PID:g61991
C;Superfamily: parainfluenza virus RNA-directed RNA polymerase

Query Match 47.6%; Score 39; DB 2; Length 2262;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy       2 ARTLNNRYTGPYTF 15
           |||||  |  |: |
Db     685 ARTLNRMYGVPHLF 698
```

RESULT 7
S43577
C28A5.4 protein (clone C28A5) - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 17-Oct-1997

```
C;Accession: S43577
R;Palmer, S.
submitted to the EMBL Data Library, April 1994
A;Reference number: S43577
A;Accession: S43577
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-282 <PAL>
A;Cross-references: EMBL:Z32680
C;Genetics:
A;Introns: 77/1; 120/2; 145/3; 225/3
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;103-168/Domain: homeobox homology #status atypical <HOX>


  Query Match             47.6%;  Score 39;  DB 2;  Length 282;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 TLNNRYTGPYTF 15
           ||| :|  || :
Db     227 TLNEQYASPYLY 238


RESULT   8
T01862
hypothetical protein T7M24.4 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C;Accession: T01862
R;Harmon, G.; Langston, Y.; Stoneking, T.; Drone, K.; Ames, M.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of Arabidopsis thaliana T7M24.
A;Reference number: Z14448
A;Accession: T01862
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-973 <HAR>
A;Cross-references: EMBL:AF077408; NID:g3319359; PID:g3319363
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 257/1; 284/1; 378/3; 689/3; 744/3
A;Note: T7M24.4


  Query Match             47.6%;  Score 39;  DB 2;  Length 973;
  Best Local Similarity   53.8%;  Pred. No. 73;
  Matches    7;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SARTLNNRYTGPY 13
           | : |: || |||
Db     826 SRKKLSPRYVGPY 838


RESULT   9
S25196
mitochondrial import protein MPI1 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YI3299.09; protein YIL022w
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 07-May-1999
C;Accession: S25196; S49960; S35357
R;Maarse, A.C.; Blom, J.; Grivell, L.A.; Meijer, M.
EMBO J. 11, 3619-3628, 1992
A;Title: MPI1, an essential gene encoding a mitochondrial membrane protein, is possibly
A;Reference number: S25196; MUID:93010953
A;Accession: S25196
A;Molecule type: DNA
A;Residues: 1-431 <MAA>
A;Cross-references: EMBL:X67276; NID:g3966; PID:g3967
R;Skelton, J.; Churcher, C.
submitted to the EMBL Data Library, December 1994
A;Reference number: S49951
A;Accession: S49960
A;Molecule type: DNA
A;Residues: 1-431 <SKE>
A;Cross-references: EMBL:Z46881; NID:g599967; PID:g599977; MIPS:YIL022w
R;Horst, M.; Jenoe, P.; Kronidou, N.G.; Bolliger, L.; Oppliger, W.; Scherer, P.; Mann
EMBO J. 12, 3035-3041, 1993
A;Title: Protein import into yeast mitochondria: the inner membrane import site prote
A;Reference number: S35357; MUID:93345448
A;Accession: S35357
A;Molecule type: protein
A;Residues: 82-89;112-122;221-228;295-306;316-327;356-370 <HOR>
C;Genetics:
A;Gene: SGD:TIM44; MPI1
A;Cross-references: MIPS:YIL022w; SGD:S0001284
A;Map position: 9L
A;Genome: nuclear
C;Keywords: membrane protein; mitochondrion


  Query Match             47.6%;  Score 39;  DB 2;  Length 431;
  Best Local Similarity   57.1%;  Pred. No. 29;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 SARTLNNRYTGPYT 14
           |:|||  ||   ||
Db      13 SSRTLTARYRSQYT 26


RESULT  10
S69044
hypothetical protein YPL141c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein LPI5
C;Species: Saccharomyces cerevisiae
C;Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 24-Sep-1999
C;Accession: S69044
R;Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S69040
A;Accession: S69044
A;Molecule type: DNA
A;Residues: 1-865 <HAL>
A;Cross-references: EMBL:U43703; NID:g1244769; PIDN:AAB68219.1; PID:g1244774; MIPS:YP
C;Genetics:
A;Map position: 16L
A;Note: YPL141c
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: ATP
F;39-313/Domain: protein kinase homology <KIN>
F;47-55/Region: protein kinase ATP-binding motif


  Query Match             46.3%;  Score 38;  DB 2;  Length 865;
  Best Local Similarity   53.8%;  Pred. No. 96;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SARTLNNRYTGPY 13
           |::||| | | |:
Db     575 SSQTLNQRDTSPF 587


RESULT  11
I49279
sauvagine/corticotropin-releasing factor receptor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 26-Aug-1999
C;Accession: I49279
R;Kishimoto, T.; Pearse, R.V.
Proc. Natl. Acad. Sci. U.S.A. 92, 1108-1112, 1995
A;Title: A sauvagine/corticotropin-releasing factor receptor expressed in heart and s
A;Reference number: I49279; MUID:95166778
A;Accession: I49279
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-431 <KIS>
A;Cross-references: EMBL:U21729; NID:g717137; PID:g717138
C;Superfamily: glucagon receptor

  Query Match             46.3%;  Score 38;  DB 2;  Length 431;
  Best Local Similarity   46.2%;  Pred. No. 44;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 RTLNNRYTGPYTF 15
          ||    ::||||:
Db     47 RTTIGNFSGPYTY 59


RESULT  12
F71151
probable glucose-1-phosphate thymidylyltransferase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 29-Sep-1999
C;Accession: F71151
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: F71151
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-356 <KAW>
A;Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29499.1; PID:d1030442; PID:g32568
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0413
C;Superfamily: mannose-1-phosphate guanylyltransferase

  Query Match             46.3%;  Score 38;  DB 2;  Length 356;
  Best Local Similarity   60.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 LNNRYTGPYT 14
          : | | ||||
Db    284 IRNAYIGPYT 293


RESULT  13
C70912
hypothetical protein Rv0043c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C;Accession: C70912
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: C70912
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-244 <COL>
A;Cross-references: GB:Z80775; GB:AL123456; NID:g3250715; PID:e268216; PID:g1568582
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0043c

  Query Match             46.3%;  Score 38;  DB 2;  Length 244;
  Best Local Similarity   60.0%;  Pred. No. 23;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 RTLNNRYTGP 12
          ||:|: | ||
Db    134 RTVNDEYAGP 143


RESULT  14
B70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C;Accession: B70931
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: B70931
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-423 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PID:e1254612; PID:g292448
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: PPE

  Query Match             46.3%;  Score 38;  DB 2;  Length 423;
  Best Local Similarity   63.6%;  Pred. No. 43;
  Matches    7;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      5 LNNRYTGPYTF 15
          |    ||||||
Db    241 LGGAVTGPYTF 251


RESULT  15
E69312
nitrate reductase gamma subunit homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C;Accession: E69312
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343
A;Accession: E69312
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-332 <KLE>
A;Cross-references: GB:AE001069; GB:AE000782; NID:g2689392; PID:g2650124; TIGR:AF0501

  Query Match             46.3%;  Score 38;  DB 2;  Length 332;
  Best Local Similarity   46.2%;  Pred. No. 33;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 RTLNNRYTGPYTF 15
          ||: :|:  |||:
Db     53 RTIYDRFDSPYTW 65


Search completed: January  7, 2000, 08:52:46
Job time: 2276 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:40 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-6
Perfect score: 82
Sequence: 1 SARTLNNRYTGPYTF 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	100.0	999	1	DSG3_HUMAN	P32926 homo sapien
2	41	50.0	391	1	GA6B_XENLA	P70005 xenopus lae
3	39	47.6	245	1	GVPL_ANAFL	P55149 anabaena fl
4	39	47.6	273	1	HMD1_CAEEL	Q18273 caenorhabdi
5	39	47.6	431	1	IM44_YEAST	Q01852 saccharomyc
6	39	47.6	2261	1	RRPL_MUMPM	P30929 mumps virus
7	39	47.6	2262	1	RRPL_PI2HT	P26676 human parai
8	38	46.3	431	1	CRF2_MOUSE	Q60748 mus musculu
9	38	46.3	1111	1	SYLC_SCHPO	Q10490 schizosacch
10	38	46.3	244	1	Y0H6_MYCTU	P71700 mycobacteri
11	37.5	45.7	406	1	YDJ1_SCHPO	P87049 schizosacch
12	37	45.1	387	1	ASPP_AEDAE	Q03168 aedes aegyp
13	37	45.1	413	1	ESC1_SCHPO	Q04635 schizosacch
14	37	45.1	643	1	SYRM_YEAST	P38714 saccharomyc
15	37	45.1	302	1	YDAK_ECOLI	P77744 escherichia
16	36	43.9	1417	1	BLM_HUMAN	P54132 homo sapien
17	36	43.9	504	1	C6B2_HELAM	Q27664 helicoverpa
18	36	43.9	362	1	COA1_POVBA	P14996 polyomaviru
19	36	43.9	372	1	COA1_POVMK	P24595 mouse polyo
20	36	43.9	340	1	COND_NEUCR	P19463 neurospora
21	36	43.9	1237	1	POL4_DROME	P10394 drosophila
22	36	43.9	3898	1	POLG_HCVB	P21530 hog cholera
23	36	43.9	2269	1	RRPL_SV41	P35341 simian viru
24	36	43.9	334	1	VG08_BPT4	P19062 bacteriopha
25	35	42.7	279	1	ASF1_YEAST	P32447 saccharomyc
26	35	42.7	383	1	COA1_POVM3	P03091 mouse polyo
27	35	42.7	382	1	COA1_POVMA	P03090 mouse polyo
28	35	42.7	383	1	COA1_POVMC	P12907 mouse polyo
29	35	42.7	383	1	COA1_POVMP	P49302 mouse polyo
30	35	42.7	325	1	KC1A_BOVIN	P35506 bos taurus
31	35	42.7	337	1	KC1A_CHICK	P70065 gallus gall
32	35	42.7	337	1	KC1A_HUMAN	P48729 homo sapien
33	35	42.7	325	1	KC1A_RAT	P97633 rattus norv
34	35	42.7	108	1	KV1E_HUMAN	P01597 homo sapien
35	35	42.7	469	1	NIFD_METTH	O27605 methanobact
36	35	42.7	4427	1	PKSL_BACSU	Q05470 bacillus su
37	35	42.7	120	1	POL1_MOUSE	P10400 mus musculu
38	35	42.7	390	1	POL3_MOUSE	P11367 mus musculu
39	35	42.7	843	1	POL_MLVAK	P03357 akr murine
40	35	42.7	1196	1	POL_MLVAV	P03356 akv murine
41	35	42.7	282	1	POL_MLVCB	P08361 cas-br-e mu
42	35	42.7	1204	1	POL_MLVF5	P26810 friend muri
43	35	42.7	1204	1	POL_MLVFF	P26809 friend muri
44	35	42.7	1204	1	POL_MLVFP	P26808 friend muri
45	35	42.7	228	1	YMCC_EMENI	P03883 emericella

ALIGNMENTS

RESULT 1
DSG3_HUMAN
ID DSG3_HUMAN STANDARD; PRT; 999 AA.
AC P32926;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92069753.
RA AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion.";
RL Cell 67:869-877(1991).
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC ----------

CC ----------
DR EMBL; M76482; AAA60230.1; -.
DR PIR; A41088; IJHUG3.
DR HSSP; P09803; 1EDH.
DR MIM; 169615; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 999 DESMOGLEIN 3.
FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 616 640 POTENTIAL.
FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 268 CADHERIN 2.
FT REPEAT 269 383 CADHERIN 3.
FT REPEAT 386 499 CADHERIN 4.
FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
FT CARBOHYD 110 110 POTENTIAL.

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match             100.0%;  Score 82;  DB 1;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SARTLNNRYTGPYTF 15
          |||||||||||||||
Db    512 SARTLNNRYTGPYTF 526


RESULT   2
GA6B_XENLA
ID   GA6B_XENLA     STANDARD;      PRT;   391 AA.
AC   P70005;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRANSCRIPTION FACTOR XGATA-6B (GATA BINDING FACTOR-6B).
GN   GATA-6A OR GATA-6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 97180937.
RA   GOVE C.D., WALMSLEY M., NIJJAR S., BERTWISTLE D., GUILLE M.,
RA   PARTINGTON G., BOMFORD A., PATIENT R.;
RT   "Over-expression of GATA-6 in Xenopus embryos blocks differentiation
RT   of heart precursors.";
RL   EMBO J. 16:355-368(1997).
CC   -!- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE
CC       CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES
CC       THE EXPRESSION OF CARDIAC MHC-ALPHA IN VIVO.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y08865; CAA70088.1; -.
DR   HSSP; P17678; 1GAU.
DR   PFAM; PF00320; GATA; 2.
DR   PROSITE; PS00344; GATA_ZN_FINGER; 2.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW   Nuclear protein.
FT   ZN_FING     182    206       GATA-TYPE.
FT   ZN_FING     236    260       GATA-TYPE.
FT   DOMAIN       70     79       POLY-SER.
FT   DOMAIN      241    245       POLY-THR.
FT   DOMAIN      293    298       POLY-SER.
SQ   SEQUENCE   391 AA;  41503 MW;  221DD428 CRC32;


  Query Match             50.0%;  Score 41;  DB 1;  Length 391;
  Best Local Similarity   61.5%;  Pred. No. 5.5;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 ARTLNNRYTGPYT 14
          || ||  |  |||
Db    120 ARPLNGSYGSPYT 132


RESULT   3
GVPL_ANAFL
ID   GVPL_ANAFL     STANDARD;      PRT;   245 AA.
AC   P55149;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   GVPF/L PROTEIN.
GN   GVPFL.
OS   Anabaena flos-aquae.
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 1403/13F;
RX   MEDLINE; 97217685.
RA   KINSMAN R., HAYES P.K.;
RT   "Genes encoding proteins homologous to halobacterial Gvps N, J, K, F
RT   & L are located downstream of gvpC in the cyanobacterium Anabaena
RT   flos-aquae.";
RL   DNA Seq. 7:97-106(1997).
CC   -!- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS
CC       VESICLE SYNTHESIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17109; AAA58714.1; -.
KW   Gas vesicle.
SQ   SEQUENCE   245 AA;  28676 MW;  F468D20E CRC32;


  Query Match             47.6%;  Score 39;  DB 1;  Length 245;
  Best Local Similarity   66.7%;  Pred. No. 7.3;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 NRYTGPYTF 15
          | :| ||||
Db    232 NNFTAPYTF 240


RESULT   4
HMD1_CAEEL
ID   HMD1_CAEEL     STANDARD;      PRT;   273 AA.
AC   Q18273;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PUTATIVE HOMEOBOX PROTEIN C28A5.4.
GN   C28A5.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   PALMER S.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEOBOX
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Z32680; CAA83601.1; -.
DR   HSSP; P02836; 1HDD.
DR   WORMPEP; C28A5.4; CE05325.
DR   PFAM; PF00046; homeobox; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Hypothetical protein; Homeobox; DNA-binding; Nuclear protein.
FT   DNA_BIND    102    161       HOMEOBOX.
SQ   SEQUENCE   273 AA;  30163 MW;  1665A6DF CRC32;


  Query Match             47.6%;  Score 39;  DB 1;  Length 273;
  Best Local Similarity   50.0%;  Pred. No. 8.3;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      4 TLNNRYTGPYTF 15
          ||| :|  || :
Db    218 TLNEQYASPYLY 229


RESULT   5
IM44_YEAST
ID   IM44_YEAST     STANDARD;      PRT;   431 AA.
AC   Q01852;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
DE   PRECURSOR (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 1) (INNER MEMBRANE
DE   IMPORT SITE PROTEIN 45) (ISP45) (MEMBRANE IMPORT MACHINERY PROTEIN
DE   MIM44).
GN   TIM44 OR MPI1 OR MIM44 OR ISP45 OR YIL022W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RX   MEDLINE; 93010953.
RA   MAARSE A.C., BLOM J., GRIVELL L.A., MEIJER M.;
RT   "MPI1, an essential gene encoding a mitochondrial membrane protein,
RT   is possibly involved in protein import into yeast mitochondria.";
RL   EMBO J. 11:3619-3628(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA   CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA   GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA   LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA   RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA   WALSH S.V., WHITEHEAD S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX   MEDLINE; 93345448.
RA   HORST M., JENOE P., KRONIDOU N.G., BOLLIGER L., OPPLIGER W.,
RA   SCHERER P., MANNING-KRIEG U., JASCUR T., SCHATZ G.;
RT   "Protein import into yeast mitochondria: the inner membrane import
RT   site protein ISP45 is the MPI1 gene product.";
RL   EMBO J. 12:3035-3041(1993).
CC   -!- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
CC       PROBABLY INVOLVED IN TRANSLOCATION ACROSS THE INNER MEMBRANE.
CC       AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
CC       PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE
CC       (MGE1) TO DRIVE PROTEIN TRANSLOCATION INTO THE MATRIX USING ATP
CC       AS AN ENERGY SOURCE.
CC   -!- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC       LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Z46881; CAA86970.1; -.
DR   EMBL; X67276; CAA47693.1; -.
DR   PIR; S25196; S25196.
DR   SGD; L0001138; MPI1.
KW   Mitochondrion; Inner membrane; Transport; Protein transport;
KW   Translocation; Transit peptide; ATP-binding.
FT   TRANSIT       1      ?       MITOCHONDRION.
FT   CHAIN         ?    431       MITOCHONDRIAL IMPORT INNER MEMBRANE
FT                                TRANSLOCASE SUBUNIT TIM44.
FT   NP_BIND     101    108       ATP (POTENTIAL).
SQ   SEQUENCE   431 AA;  48854 MW;  002E0771 CRC32;


  Query Match             47.6%;  Score 39;  DB 1;  Length 431;
  Best Local Similarity   57.1%;  Pred. No. 14;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 SARTLNNRYTGPYT 14
          |:|||  ||   ||
Db     13 SSRTLTARYRSQYT 26


RESULT   6
RRPL_MUMPM
ID   RRPL_MUMPM     STANDARD;      PRT;  2261 AA.
AC   P30929;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Mumps virus (strain Miyahara vaccine).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92263804.
RA   OKAZAKI K., TANABAYASHI K., TAKEUCHI K., HISHIYAMA M., OKAZAKI K.,
RA   YAMADA A.;
RT   "Molecular cloning and sequence analysis of the mumps virus gene
RT   encoding the L protein and the trailer sequence.";
RL   Virology 188:926-930(1992).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; D10575; BAA01432.1; -.
DR   PIR; A42548; A42548.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2261 AA;  256572 MW;  A6CC5C12 CRC32;
```

```
  Query Match             47.6%;  Score 39;  DB 1;  Length 2261;
  Best Local Similarity   57.1%;  Pred. No. 91;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ARTLNNRYTGPYTF 15
           |||||: |  |: |
Db     686 ARTLNSMYGIPHLF 699


RESULT   7
RRPL_PI2HT
ID   RRPL_PI2HT     STANDARD;      PRT;  2262 AA.
AC   P26676;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Human parainfluenza 2 virus (strain Toshiba) (PIV-2).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91252221.
RA   KAWANO M., OKAMOTO K., BANDO H., KONDO K., TSURUDOME M., KOMADA H.,
RA   NISHIO M., ITO Y.;
RT   "Characterizations of the human parainfluenza type 2 virus gene
RT   encoding the L protein and the intergenic sequences.";
RL   Nucleic Acids Res. 19:2739-2746(1991).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57559; CAA40788.1; -.
DR   PIR; S16664; S16664.
DR   PFAM; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2262 AA;  256380 MW;  2EF04669 CRC32;


  Query Match             47.6%;  Score 39;  DB 1;  Length 2262;
  Best Local Similarity   57.1%;  Pred. No. 91;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 ARTLNNRYTGPYTF 15
           |||||  |  |: |
Db     685 ARTLNRMYGVPHLF 698


RESULT   8
CRF2_MOUSE
ID   CRF2_MOUSE     STANDARD;      PRT;   431 AA.
AC   Q60748; Q60808; Q60783;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CORTICOTROPIN RELEASING FACTOR RECEPTOR 2 PRECURSOR (CRF-R) (CRF2)
DE   (CRF-RB) (CRH-R2).
GN   CRHR2 OR CRF2R.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE; 95224061.
RA   PERRIN M., DONALDSON C., CHEN R., BLOUNT A., BERGGREN T.,
RA   BILEZIKJIAN L., SAWCHENKO P., VALE W.;
RT   "Identification of a second corticotropin-releasing factor receptor
RT   gene and characterization of a cDNA expressed in heart.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:2969-2973(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=HEART;
RX   MEDLINE; 95166778.
RA   KISHIMOTO T., PEARSE R.V. II, LIN C.R., ROSENFELD M.G.;
RT   "A sauvagine/corticotropin-releasing factor receptor expressed in
RT   heart and skeletal muscle.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:1108-1112(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=HEART;
RX   MEDLINE; 96015396.
RA   STENZEL P., KESTERSON R., YEUNG W., CONE R.D., RITTENBERG M.B.,
RA   STENZEL-POORE M.P.;
RT   "Identification of a novel murine receptor for
RT   corticotropin-releasing hormone expressed in the heart.";
RL   Mol. Endocrinol. 9:637-645(1995).
CC   -!- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC       SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
CC       MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE HEART. ALSO EXPRESSED
CC       IN LUNGS, SKELETAL MUSCLE, GASTROINTESTINAL TRACT, EPIDIDYMIS, AND
CC       BRAIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17858; AAA68026.1; -.
DR   EMBL; U21729; AAC52174.1; -.
DR   EMBL; U19939; AAC52243.1; -.
DR   GCRDB; GCR_1705; -.
DR   GCRDB; GCR_1711; -.
DR   MGD; MGI:894312; CRHR2.
DR   PFAM; PF00002; 7tm_2; 1.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL       1    24       POTENTIAL.
FT   CHAIN       25   431       CORTICOTROPIN RELEASING FACTOR RECEPTOR
FT                              2.
FT   DOMAIN      25   138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   139   159       1 (POTENTIAL).
FT   DOMAIN     160   168       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   169   188       2 (POTENTIAL).
FT   DOMAIN     189   205       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   206   229       3 (POTENTIAL).
FT   DOMAIN     230   243       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   244   265       4 (POTENTIAL).
FT   DOMAIN     266   284       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   285   307       5 (POTENTIAL).
FT   DOMAIN     308   330       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   331   350       6 (POTENTIAL).
FT   DOMAIN     351   365       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   366   385       7 (POTENTIAL).
FT   DOMAIN     386   431       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    52    52       POTENTIAL.
FT   CARBOHYD    61    61       POTENTIAL.
```

```
Qy       1 QSGTMRTRHSTGGTN 15
           | | : | |:| | |
Db     387 QDGALNTGHTTSGVN 401


RESULT   4
HEMA_PI3B
ID   HEMA_PI3B      STANDARD;      PRT;   572 AA.
AC   P06167;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).
GN   HN.
OS   Bovine parainfluenza 3 virus.
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=910N;
RX   MEDLINE; 87174819.
RA   SUZU S., SAKAI Y., SHIODA T., SHIBUTA H.;
RT   "Nucleotide sequence of the bovine parainfluenza 3 virus genome: the
RT   genes of the F and HN glycoproteins.";
RL   Nucleic Acids Res. 15:2945-2958(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC       TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEUROAMINIDASE
CC       ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC       THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
CC       GLYCOPROTEINS.
CC   -!- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
CC       N-TERMINAL HYDROPHOBIC SEQUENCE.
CC   -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
CC       NEURAMINIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00114; CAA68298.1; -.
DR   EMBL; D84095; BAA12218.1; -.
DR   PIR; B27218; HNNZB3.
DR   PFAM; PF00423; HN; 1.
KW   Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;
KW   Transmembrane.
FT   DOMAIN        1     30       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     31     53       POTENTIAL.
FT   DOMAIN       54    572       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    308    308       POTENTIAL.
FT   CARBOHYD    351    351       POTENTIAL.
FT   CARBOHYD    448    448       POTENTIAL.
FT   CARBOHYD    523    523       POTENTIAL.
FT   CARBOHYD    570    570       POTENTIAL.
SQ   SEQUENCE   572 AA;  64590 MW;  876CFFE6 CRC32;


  Query Match             51.2%;  Score 41;  DB 1;  Length 572;
  Best Local Similarity   61.5%;  Pred. No. 7.3;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 GTMRTRHSTGGTN 15
           || | |||:  ||
Db      18 GTTRDRHSSKATN 30


RESULT   5
AMY_BACLI
ID   AMY_BACLI      STANDARD;      PRT;   512 AA.
AC   P06278;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ALPHA-AMYLASE PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DE   GLUCANOHYDROLASE).
GN   AMYS OR AMYL.
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27811;
RX   MEDLINE; 86111694.
RA   YUUKI T., NOMURA T., TEZUKA H., TSUBOI A., YAMAGATA H.,
RA   TSUKAGOSHI N., UDAKA S.;
RT   "Complete nucleotide sequence of a gene coding for heat- and
RT   pH-stable alpha-amylase of Bacillus licheniformis: comparison of the
RT   amino acid sequences of three bacterial liquefying alpha-amylases
RT   deduced from the DNA sequences.";
RL   J. Biochem. 98:1147-1156(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86195857.
RA   GRAY G.L., MAINZER S.E., REY M.W., LAMSA M.H., KINDLE K.L.,
RA   CARMONA C., REQUADT C.;
RT   "Structural genes encoding the thermophilic alpha-amylases of
RT   Bacillus stearothermophilus and Bacillus licheniformis.";
RL   J. Bacteriol. 166:635-643(1986).
RN   [3]
RP   SEQUENCE OF 1-104 FROM N.A.
RX   MEDLINE; 84185455.
RA   STEPHENS M.A., ORTLEPP S.A., OLLINGTON J.F., MCCONNELL D.J.;
RT   "Nucleotide sequence of the 5' region of the Bacillus licheniformis
RT   alpha-amylase gene: comparison with the B. amyloliquefaciens gene.";
RL   J. Bacteriol. 158:369-372(1984).
RN   [4]
RP   SEQUENCE OF 1-29 FROM N.A.
RX   MEDLINE; 89213924.
RA   LAOIDE B.M., CHAMBLISS G.H., MCCONNELL D.J.;
RT   "Bacillus licheniformis alpha-amylase gene, amyL, is subject to
RT   promoter-independent catabolite repression in Bacillus subtilis.";
RL   J. Bacteriol. 171:2435-2442(1989).
RN   [5]
RP   SEQUENCE OF 30-47.
RX   MEDLINE; 82098050.
RA   KUHN H., FIETZEK P.P., LAMPEN J.O.;
RT   "N-terminal amino acid sequence of Bacillus licheniformis
RT   alpha-amylase: comparison with Bacillus amyloliquefaciens and
RT   Bacillus subtilis Enzymes.";
RL   J. Bacteriol. 149:372-373(1982).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC   STRAIN=ATCC 27811;
RX   MEDLINE; 95182462.
RA   MACHIUS M., WIEGAND G., HUBER R.;
RT   "Crystal structure of calcium-depleted Bacillus licheniformis alpha-
RT   amylase at 2.2-A resolution.";
RL   J. Mol. Biol. 246:545-559(1995).
CC   -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC       LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC       KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03236; CAA26981.1; -.
```

```
DR   EMBL; M38570; AAA22226.1; -.
DR   EMBL; M13256; AAA22240.1; -.
DR   EMBL; K01984; AAA22193.1; -.
DR   EMBL; M26412; AAA22237.1; -.
DR   EMBL; A17930; CAA01355.1; -.
DR   PIR; A00844; ALBSL.
DR   PIR; B24549; B24549.
DR   PIR; A26151; A26151.
DR   PDB; 1BPL; 17-AUG-96.
DR   PDB; 1VJS; 12-MAR-97.
DR   PFAM; PF00128; alpha-amylase; 1.
KW   Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; 3D-structure.
FT   SIGNAL       1     29
FT   CHAIN       30    512       ALPHA-AMYLASE.
FT   ACT_SITE    260    260       BY SIMILARITY.
FT   ACT_SITE    264    264       BY SIMILARITY.
FT   ACT_SITE    357    357       BY SIMILARITY.
FT   CONFLICT     38     38       Q -> Y (IN REF. 5).
FT   CONFLICT    163    163       R -> L (IN REF. 2).
FT   CONFLICT    339    339       S -> G (IN REF. 2).
FT   CONFLICT    349    349       A -> S (IN REF. 2).
SQ   SEQUENCE   512 AA;  58549 MW;  4B5D334D CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 512;
  Best Local Similarity   53.8%;  Pred. No. 9.7;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 QSGTMRTRHSTGG 13
          | ||:||:: | |
Db     98 QKGTVRTKYGTKG 110


RESULT   6
CATB_PSEAE
ID   CATB_PSEAE     STANDARD;      PRT;   513 AA.
AC   Q59635;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CATALASE PRECURSOR (EC 1.11.1.6) (PARAQUAT INDUCIBLE CATALASE ISOZYME
DE   B).
GN   KATB.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRD2;
RX   MEDLINE; 96062238.
RA   BROWN S.M., HOWELL M.L., VASIL M.L., ANDERSON A.J., HASSETT D.J.;
RT   "Cloning and characterization of the katB gene of Pseudomonas
RT   aeruginosa encoding a hydrogen peroxide-inducible catalase:
RT   purification of KatB, cellular localization, and demonstration that
RT   it is essential for optimal resistance to hydrogen peroxide.";
RL   J. Bacteriol. 177:6536-6544(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRD1;
RA   HOWELL M.L., HEUR M., KLOTZ M.G., HASSETT D.J.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
CC       TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: HEME GROUP.
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U34896; AAA79046.1; -.
DR   EMBL; U89384; AAB49463.1; -.
DR   HSSP; P00432; 7CAT.
DR   PFAM; PF00199; catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW   Periplasmic; Signal.
FT   SIGNAL       1     30       POTENTIAL.
FT   CHAIN       31    513       CATALASE.
FT   ACT_SITE     81     81       BY SIMILARITY.
FT   ACT_SITE    153    153       BY SIMILARITY.
FT   BINDING     361    361       PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ   SEQUENCE   513 AA;  57131 MW;  D18FF8ED CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 513;
  Best Local Similarity   46.7%;  Pred. No. 9.7;
  Matches    7;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 QSGTMRTRHSTGGTN 15
          | | :   ||| | |
Db    390 QDGALNAGHSTSGVN 404


RESULT   7
FILA_HUMAN
ID   FILA_HUMAN     STANDARD;      PRT;   416 AA.
AC   P20930;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FILAGGRIN PRECURSOR (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89296901.
RA   MCKINLEY-GRANT L.J., IDLER W.W., BERNSTEIN I.A., PARRY D.A.D.,
RA   CANNIZZARO L., CROCE C.M., HUEBNER K., LESSIN S.R., STEINERT P.M.;
RT   "Characterization of a cDNA clone encoding human filaggrin and
RT   localization of the gene to chromosome region 1q21.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
CC   -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC       DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC       TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC   -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC       HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC       OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC       PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC       DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M24355; AAA52454.1; -.
DR   PIR; A32947; A32947.
DR   MIM; 135940; -.
KW   Phosphorylation; Polyprotein; Developmental protein.
FT   NON_TER      1      1
SQ   SEQUENCE   416 AA;  44105 MW;  72D2B913 CRC32;
```

```
  Query Match             50.0%;  Score 40;  DB 1;  Length 416;
  Best Local Similarity   61.5%;  Pred. No. 7.8;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 QSGTMRTRHSTGG 13
           ||||  |: |:||
Db     237 QSGTRHTQTSSGG 249


RESULT   8
FYN_CHICK
ID   FYN_CHICK      STANDARD;      PRT;   533 AA.
AC   Q05876;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN   FYN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=MUSCLE;
RX   MEDLINE; 93205395.
RA   SUDOL M., GREULICH H., NEWMAN L., SARKAR A., SUKEGAWA J., YAMAMOTO T.;
RT   "A novel Yes-related kinase, Yrk, is expressed at elevated levels in
RT   neural and hematopoietic tissues.";
RL   Oncogene 8:823-831(1993).
CC   -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- TISSUE SPECIFICITY: THYMUS AND SPLEEN.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52841; CAA37025.1; -.
DR   PIR; S36351; S36351.
DR   HSSP; P06241; 1FYN.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW   ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;
KW   Lipoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   LIPID         2      2       PALMITATE (BY SIMILARITY).
FT   LIPID         5      5       PALMITATE (BY SIMILARITY).
FT   DOMAIN       81    142       SH3.
FT   DOMAIN      148    245       SH2.
FT   DOMAIN      267    520       PROTEIN KINASE.
FT   MOD_RES      11     11       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   NP_BIND     273    281       ATP (BY SIMILARITY).
FT   BINDING     295    295       ATP (BY SIMILARITY).
FT   ACT_SITE    386    386       BY SIMILARITY.
FT   MOD_RES     416    416       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     527    527       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   533 AA;  60134 MW;  05970E9A CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 533;
  Best Local Similarity   61.5%;  Pred. No. 10;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 SGTMRTRHSTGGT 14
           :||:|||  || |
Db      72 TGTLRTRGGTGVT 84


RESULT   9
FYN_HUMAN
ID   FYN_HUMAN      STANDARD;      PRT;   536 AA.
AC   P06241;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN)
DE   (SYN) (SLK).
GN   FYN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86287278.
RA   SEMBA K., NISHIZAWA M., MIYAJIMA N., YOSHIDA M.C., SUKEGAWA J.,
RA   YAMANASHI Y., SASAKI M., YAMAMOTO T., TOYOSHIMA K.;
RT   "Yes-related protooncogene, syn, belongs to the protein-tyrosine
RT   kinase family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5459-5463(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87089775.
RA   KAWAKAMI T., PENNINGTON C.Y., ROBBINS K.C.;
RT   "Isolation and oncogenic potential of a novel human src-like gene.";
RL   Mol. Cell. Biol. 6:4195-4201(1986).
RN   [3]
RP   MYRISTOYLATION, AND PHOSPHORYLATION AT TYR-530.
RX   MEDLINE; 91016431.
RA   PETERS D.J., MCGREW B.R., PERRON D.C., LIPTAK L.M., LAUDANO A.P.;
RT   "In vivo phosphorylation and membrane association of the fyn proto-
RT   oncogene product in IM-9 human lymphoblasts.";
RL   Oncogene 5:1313-1319(1990).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF SH3 DOMAIN.
RX   MEDLINE; 93327750.
RA   NOBLE M.E.M., MUSACCHIO A., SARASTE M., COURTNEIDGE S.A.,
RA   WIERENGA R.K.;
RT   "Crystal structure of the SH3 domain in human Fyn; comparison of the
RT   three-dimensional structures of SH3 domains in tyrosine kinases and
RT   spectrin.";
RL   EMBO J. 12:2617-2624(1993).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 80-141.
RX   MEDLINE; 95393198.
RA   MUSACCHIO A., SARASTE M., WILMANNS M.;
RT   "High-resolution crystal structures of tyrosine kinase SH3 domains
RT   complexed with proline-rich peptides.";
RL   Nat. Struct. Biol. 1:546-551(1994).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 84-140 IN COMPLEX WITH NEF.
RX   MEDLINE; 96279837.
RA   LEE C.H., SAKSELA K., MIRZA U.A., CHAIT B.T., KURIYAN J.;
RT   "Crystal structure of the conserved core of HIV-1 Nef complexed with
RT   a Src family SH3 domain.";
RL   Cell 85:931-942(1996).
RN   [7]
RP   STRUCTURE BY NMR OF SH3 DOMAIN.
```

```
RX   MEDLINE; 96399716.
RA   MORTON C.J., PUGH D.J.R., BROWN E.L.J., KAHMANN J.D., RENZONI D.A.C.,
RA   CAMPBELL I.D.;
RT   "Solution structure and peptide binding of the SH3 domain from human
RT   Fyn.";
RL   Structure 4:705-714(1996).
RN   [8]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 97121261.
RA   RENZONI D.A., PUGH D.J., SILIGARDI G., DAS P., MORTON C.J., ROSSI C.,
RA   WATERFIELD M.D., CAMPBELL I.D., LADBURY J.E.;
RT   "Structural and thermodynamic characterization of the interaction of
RT   the SH3 domain from Fyn with the proline-rich binding site on the p85
RT   subunit of PI3-kinase.";
RL   Biochemistry 35:15646-15653(1996).
RN   [9]
RP   STRUCTURE BY NMR OF SH2 DOMAIN.
RX   MEDLINE; 98035454.
RA   MULHERN T.D., SHAW G.L., MORTON C.J., DAY A.J., CAMPBELL I.D.;
RT   "The SH2 domain from the tyrosine kinase Fyn in complex with a
RT   phosphotyrosyl peptide reveals insights into domain stability and
RT   binding specificity.";
RL   Structure 5:1313-1323(1997).
RN   [10]
RP   BINDING OF SH3 DOMAIN TO PI 3-KINASE.
RX   MEDLINE; 93348274.
RA   PRASAD K.V., JANSSEN O., KAPELLER R., RAAB M., CANTLEY L.C.,
RA   RUDD C.E.;
RT   "Src-homology 3 domain of protein kinase p59fyn mediates binding to
RT   phosphatidylinositol 3-kinase in T cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:7366-7370(1993).
CC   -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14333; AAC08285.1; -.
DR   EMBL; M14676; AAA36615.1; -.
DR   PIR; A24314; TVHUSY.
DR   PIR; A25389; TVHUSR.
DR   PDB; 1SHF; 31-OCT-93.
DR   PDB; 1FYN; 08-NOV-96.
DR   PDB; 1NYF; 08-NOV-96.
DR   PDB; 1NYG; 08-NOV-96.
DR   PDB; 1EFN; 11-JAN-97.
DR   PDB; 1A0N; 25-FEB-98.
DR   PDB; 1AOT; 14-JAN-98.
DR   PDB; 1AOU; 14-JAN-98.
DR   PDB; 1AZG; 25-FEB-98.
DR   MIM; 137025; -.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW   ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;
KW   Lipoprotein; 3D-structure.
FT   INIT_MET      0      0
FT   LIPID         1      1       MYRISTATE.
FT   LIPID         2      2       PALMITATE (BY SIMILARITY).
FT   LIPID         5      5       PALMITATE (BY SIMILARITY).
FT   DOMAIN       81    142       SH3.
FT   DOMAIN      148    245       SH2.
FT   DOMAIN      270    523       PROTEIN KINASE.
FT   MOD_RES      11     11       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   NP_BIND     276    284       ATP (BY SIMILARITY).
FT   BINDING     298    298       ATP (BY SIMILARITY).
FT   ACT_SITE    389    389       BY SIMILARITY.
FT   MOD_RES     419    419       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     530    530       PHOSPHORYLATION.
FT   CONFLICT    183    183       A -> S (IN REF. 2).
FT   CONFLICT    436    436       A -> R (IN REF. 2).
FT   STRAND       85     88
FT   STRAND       92     92
FT   STRAND       99     99
FT   STRAND      102    102
FT   TURN        104    105
FT   STRAND      107    112
FT   STRAND      118    123
FT   TURN        124    126
FT   STRAND      129    133
FT   HELIX       134    136
FT   STRAND      137    139
SQ   SEQUENCE   536 AA;  60630 MW;  01B18DD0 CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 536;
  Best Local Similarity   61.5%;  Pred. No. 10;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 SGTMRTRHSTGGT 14
           :||:|||  || |
Db      72 TGTLRTRGGTGVT 84


RESULT  10
FYN_MOUSE
ID   FYN_MOUSE      STANDARD;      PRT;   533 AA.
AC   P39688;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN   FYN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91175680.
RA   COOKE M.P., PERLMUTTER R.M.;
RT   "Expression of a novel form of the fyn proto-oncogene in
RT   hematopoietic cells.";
RL   New Biol. 1:66-74(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LEE C., KIM M.G., JEON S.H., PARK D.E., PARK S.D., SEONG R.H.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PALMITOYLATION.
RX   MEDLINE; 94019312.
RA   SHENOY-SCARIA A.M., TIMSON L.K., KWONG J., SHAW A.S., LUBLIN D.M.;
RT   "Palmitylation of an amino-terminal cysteine motif of protein
RT   tyrosine kinases p56lck and p59fyn mediates interaction with
RT   glycosyl-phosphatidylinositol-anchored proteins.";
RL   Mol. Cell. Biol. 13:6385-6392(1993).
RN   [4]
RP   PALMITOYLATION.
RX   MEDLINE; 95071286.
```

```
RA   KOEGL M., ZLATKINE P., LEY S.C., COURTNEIDGE S.A., MAGEE A.I.;
RT   "Palmitoylation of multiple Src-family kinases at a homologous N-
RT   terminal motif.";
RL   Biochem. J. 303:749-753(1994).
RN   [5]
RP   PALMITOYLATION.
RX   MEDLINE; 97345356.
RA   WOLVEN A., OKAMURA H., ROSENBLATT Y., RESH M.D.;
RT   "Palmitoylation of p59fyn is reversible and sufficient for plasma
RT   membrane association.";
RL   Mol. Biol. Cell 8:1159-1173(1997).
RN   [6]
RP   MYRISTOYLATION.
RX   MEDLINE; 96251668.
RA   GAUEN L.K.T., LINDER M.E., SHAW A.S.;
RT   "Multiple features of the p59fyn src homology 4 domain define a motif
RT   for immune-receptor tyrosine-based activation motif (ITAM) binding
RT   and for plasma membrane localization.";
RL   J. Cell Biol. 133:1007-1015(1996).
CC   -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27266; AAA37644.1; -.
DR   EMBL; U70324; AAB09568.1; -.
DR   PIR; A44991; A44991.
DR   HSSP; P06241; 1FYN.
DR   MGD; MGI:95602; FYN.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW   ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;
KW   Lipoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE.
FT   LIPID         2      2       PALMITATE.
FT   LIPID         5      5       PALMITATE.
FT   DOMAIN       81    142       SH3.
FT   DOMAIN      148    245       SH2.
FT   DOMAIN      267    520       PROTEIN KINASE.
FT   NP_BIND     273    281       ATP (BY SIMILARITY).
FT   BINDING     295    295       ATP (BY SIMILARITY).
FT   ACT_SITE    386    386       BY SIMILARITY.
FT   MOD_RES     416    416       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   533 AA;  59926 MW;  B235E6FA CRC32;
```

```
  Query Match           50.0%;  Score 40;  DB 1;  Length 533;
  Best Local Similarity  61.5%;  Pred. No. 10;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 SGTMRTRHSTGGT 14
           :||:|||  || |
Db      72 TGTLRTRGGTGVT 84
```

```
RESULT  11
FYN_XENLA
ID   FYN_XENLA      STANDARD;      PRT;   536 AA.
AC   P13406;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN   FYN.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90191723.
RA   STEELE R.E., DENG J.C., GHOSN C.R., FERO J.B.;
RT   "Structure and expression of fyn genes in Xenopus laevis.";
RL   Oncogene 5:369-376(1990).
CC   -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52188; CAA36435.1; -.
DR   EMBL; M27502; AAA49719.1; -.
DR   PIR; A43806; A43806.
DR   HSSP; P06241; 1FYN.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW   ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;
KW   Lipoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   LIPID         2      2       PALMITATE (BY SIMILARITY).
FT   LIPID         5      5       PALMITATE (BY SIMILARITY).
FT   DOMAIN       81    142       SH3.
FT   DOMAIN      148    245       SH2.
FT   DOMAIN      270    523       PROTEIN KINASE.
FT   MOD_RES      11     11       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   NP_BIND     276    284       ATP (BY SIMILARITY).
FT   BINDING     298    298       ATP (BY SIMILARITY).
FT   ACT_SITE    389    389       BY SIMILARITY.
FT   MOD_RES     419    419       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   536 AA;  60715 MW;  87236822 CRC32;
```

```
  Query Match           50.0%;  Score 40;  DB 1;  Length 536;
  Best Local Similarity  61.5%;  Pred. No. 10;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      2 SGTMRTRHSTGGT 14
          :||:|||  || |
Db     72 TGTLRTRGGTGVT 84


RESULT  12
FYN_XIPHE
ID   FYN_XIPHE      STANDARD;      PRT;   536 AA.
AC   P27446;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN   FYN.
OS   Xiphophorus helleri.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Atherinomorpha;
OC   Cyprinodontiformes; Cyprinodontoidei; Poeciliidae; Xiphophorus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIO LANCETILLA;
RX   MEDLINE; 91187435.
RA   HANNIG G., OTTILIE S., SCHARTL M.;
RT   "Conservation of structure and expression of the c-yes and fyn genes
RT   in lower vertebrates.";
RL   Oncogene 6:361-369(1991).
CC   -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC       PHOSPHATIDYLINOSITOL 3-KINASE.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54971; CAA38715.1; -.
DR   HSSP; P06241; 1AOT.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; Transferase; Tyrosine-protein kinase; Phosphorylation;
KW   ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;
KW   Lipoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   LIPID         2      2       PALMITATE (BY SIMILARITY).
FT   LIPID         5      5       PALMITATE (BY SIMILARITY).
FT   DOMAIN       81    142       SH3.
FT   DOMAIN      148    245       SH2.
FT   DOMAIN      270    523       PROTEIN KINASE.
FT   MOD_RES      11     11       PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT   NP_BIND     276    284       ATP (BY SIMILARITY).
FT   BINDING     298    298       ATP (BY SIMILARITY).
FT   ACT_SITE    389    389       BY SIMILARITY.
FT   MOD_RES     419    419       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     530    530       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   536 AA;  60316 MW;  BA03DB12 CRC32;

  Query Match           50.0%;  Score 40;  DB 1;  Length 536;
  Best Local Similarity  61.5%;  Pred. No. 10;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SGTMRTRHSTGGT 14
          :||:|||  || |
Db     72 TGTLRTRGGTGVT 84


RESULT  13
MA5D_SCHJA
ID   MA5D_SCHJA     STANDARD;      PRT;   120 AA.
AC   P13411;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   IMMUNOGENIC MIRACIDIAL ANTIGEN 5D (FRAGMENT).
GN   5D.
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PHILIPPINE;
RX   MEDLINE; 89261924.
RA   SCALLON B.J., BOGITSH B.J., CARTER C.E.;
RT   "Characterization of a large gene family in Schistosoma japonicum
RT   that encodes an immunogenic miracidial antigen.";
RL   Mol. Biochem. Parasitol. 33:105-112(1989).
CC   -!- DEVELOPMENTAL STAGE: MIRACIDIA.
CC   -!- SIMILARITY: NEARLY IDENTICAL TO IMMUNOGENIC MIRACIDIAL
CC       ANTIGENS 8I', 8I AND 8C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25214; AAA29851.1; -.
KW   Antigen; Multigene family.
FT   NON_TER       1      1
SQ   SEQUENCE   120 AA;  13541 MW;  C5BE9039 CRC32;

  Query Match           50.0%;  Score 40;  DB 1;  Length 120;
  Best Local Similarity  57.1%;  Pred. No. 2.1;
  Matches    8;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QSGTMRTRHSTGGT 14
          | |  | ||  |||
Db     77 QGGYQRNRHYGGGT 90


RESULT  14
RPB1_EUPOC
ID   RPB1_EUPOC     STANDARD;      PRT;   478 AA.
AC   P28364;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)
DE   (FRAGMENT).
GN   RPB1.
OS   Euplotes octocarinatus.
OC   Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotes.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=(68)-VIII;
RX   MEDLINE; 93027138.
```

```
RA   KAUFMANN J., KLEIN A.;
RT   "Gene dosage as a possible major determinant for equal expression
RT   levels of genes encoding RNA polymerase subunits in the hypotrichous
RT   ciliate Euplotes octocarinatus.";
RL   Nucleic Acids Res. 20:4445-4450(1992).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO
CC       TWELVE DIFFERENT POLYPEPTIDES. THIS POLYPEPTIDE IS THE LARGEST
CC       COMPONENT OF RNA POLYMERASE II.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66452; CAA47068.1; -.
DR   PIR; S33886; S33886.
DR   PFAM; PF00623; RNA_pol_A; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW   DNA-binding; Nuclear protein.
FT   ZN_FING      68     84       C2H2-TYPE (POTENTIAL).
FT   NON_TER     478    478
SQ   SEQUENCE   478 AA;  54026 MW;  9D100D91 CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 478;
  Best Local Similarity   53.8%;  Pred. No. 9;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 QSGTMRTRHSTGG 13
          |:| | :|| :||
Db    301 QAGQMVSRHKSGG 313


RESULT  15
SCYD_MAGGR
ID   SCYD_MAGGR     STANDARD;      PRT;   172 AA.
AC   P56221;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   SCYTALONE DEHYDRATASE (EC 4.2.1.94).
OS   Magnaporthe grisea (Rice blast fungus).
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Phyllachorales; Magnaporthaceae; Magnaporthe.
RN   [1]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX   MEDLINE; 95171111.
RA   LUNDQVIST T., RICE J., HODGE C.N., BASARAB G.S., PIERCE J.,
RA   LINDQVIST Y.;
RT   "Crystal structure of scytalone dehydratase -- a disease determinant
RT   of the rice pathogen, Magnaporthe grisea.";
RL   Structure 2:937-944(1994).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE; 98332516.
RA   NAKASAKO M., MOTOYAMA T., KURAHASHI Y., YAMAGUCHI I.;
RT   "Cryogenic X-ray crystal structure analysis for the complex of
RT   scytalone dehydratase of a rice blast fungus and its tight-binding
RT   inhibitor, carpropamid: the structural basis of tight-binding
RT   inhibition.";
RL   Biochemistry 37:9931-9939(1998).
CC   -!- CATALYTIC ACTIVITY: SCYTALONE = 1,3,8-TRIHYDROXYNAPHTHALENE +
CC       H(2)O.
CC   -!- PATHWAY: BIOSYNTHESIS OF FUNGAL MELANIN.
CC   -!- SUBUNIT: HOMOTRIMER.
DR   PDB; 1STD; 19-AUG-95.
DR   PDB; 2STD; 16-FEB-99.
KW   Lyase; Melanin biosynthesis; 3D-structure.
SQ   SEQUENCE   172 AA;  20250 MW;  0D93A15B CRC32;


  Query Match             50.0%;  Score 40;  DB 1;  Length 172;
  Best Local Similarity   63.6%;  Pred. No. 3.1;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 TMRTRHSTGGT 14
          |:|:|  |||
Db     80 TLRTQHFIGGT 90


Search completed: January  7, 2000, 13:25:43
Job time: 395 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:24 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-7
Perfect score: 80
Sequence: 1 QSGTMRTRHSTGGTN 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	86.2	993	11	O35902	O35902 mus musculu
2	44	55.0	421	2	O86590	O86590 streptomyce
3	43	53.8	1597	5	O61346	O61346 drosophila
4	41	51.2	572	12	Q65689	Q65689 bovine para
5	40	50.0	534	4	Q16248	Q16248 homo sapien
6	40	50.0	584	5	Q25289	Q25289 leishmania
7	40	50.0	1921	5	O01349	O01349 drosophila
8	40	50.0	537	11	Q62844	Q62844 rattus norv
9	40	50.0	69	12	P90363	P90363 human cytom
10	40	50.0	69	12	P90292	P90292 human cytom
11	40	50.0	83	12	Q68570	Q68570 human cytom
12	40	50.0	171	12	P89538	P89538 human cytom
13	40	50.0	170	12	P90392	P90392 human cytom
14	40	50.0	169	12	P90393	P90393 human cytom
15	40	50.0	172	12	P90394	P90394 human cytom
16	40	50.0	718	12	Q69174	Q69174 human cytom
17	40	50.0	905	12	Q69173	Q69173 human cytom
18	39	48.8	275	2	Q44378	Q44378 agrobacteri
19	39	48.8	1879	2	P72938	P72938 synechocyst
20	39	48.8	797	5	O44536	O44536 caenorhabdi
21	38	47.5	1089	2	O07800	O07800 mycobacteri
22	38	47.5	137	2	P74924	P74924 thermotoga
23	38	47.5	425	5	O44433	O44433 drosophila
24	38	47.5	1060	5	Q16926	Q16926 aedes aegyp
25	38	47.5	396	5	Q22250	Q22250 caenorhabdi
26	38	47.5	428	5	O96627	O96627 drosophila
27	38	47.5	1248	10	O65589	O65589 arabidopsis
28	38	47.5	859	11	O88509	O88509 mus musculu
29	38	47.5	839	11	O88510	O88510 mus musculu
30	38	47.5	776	11	O88511	O88511 mus musculu
31	38	47.5	204	12	Q77371	Q77371 human immun
32	38	47.5	469	12	P88837	P88837 influenza a
33	38	47.5	469	12	P88838	P88838 influenza a
34	38	47.5	850	12	Q97008	Q97008 human immun
35	37.5	46.9	1708	5	O16039	O16039 drosophila
36	37	46.2	136	2	O51792	O51792 alcaligenes
37	37	46.2	466	2	O88021	O88021 streptomyce
38	37	46.2	190	2	O07902	O07902 bacillus sp
39	37	46.2	744	3	Q12715	Q12715 trichoderma
40	37	46.2	700	4	P78505	P78505 homo sapien
41	37	46.2	791	4	O43849	O43849 homo sapien
42	37	46.2	782	4	O76105	O76105 homo sapien
43	37	46.2	1120	4	O95213	O95213 homo sapien
44	37	46.2	670	4	O95447	O95447 homo sapien
45	37	46.2	357	5	Q18488	Q18488 caenorhabdi

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 86.2%; Score 69; DB 11; Length 993;
Best Local Similarity 100.0%; Pred. No. 0.00051;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SGTMRTRHSTGGT 14
||||||||||||||
Db 743 SGTMRTRHSTGGT 755

RESULT 2
O86590
ID O86590 PRELIMINARY; PRT; 421 AA.
AC O86590;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SC2H4.07C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   OLIVER K., HARRIS D.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE; 97000351.
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL031514; CAA20601.1; -.
SQ   SEQUENCE   421 AA;  44174 MW;  7B444909 CRC32;


  Query Match             55.0%;  Score 44;  DB 2;  Length 421;
  Best Local Similarity   69.2%;  Pred. No. 4.4;
  Matches    9;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SGTMRTRHSTGGT 14
          ||:||||   |||
Db     48 SGSMRTRDIDGGT 60


RESULT   3
O61346
ID   O61346      PRELIMINARY;      PRT;  1597 AA.
AC   O61346;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SMALL OPTIC LOBES.
GN   SOL.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTONS;
RX   MEDLINE; 96129280.
RA   DE COUET H.G., FONG K.S., WEEDS A.G., MCLAUGHLIN P.J., MIKLOS G.L.;
RT   "Molecular and mutational analysis of a gelsolin-family member
RT   encoded by the flightless I gene of Drosophila melanogaster.";
RL   Genetics 141:1049-1059(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTONS;
RX   MEDLINE; 97289742.
RA   MIKLOS G.L., YAMAMOTO M., BURNS R.G., MALESZKA R.;
RT   "An essential cell division gene of Drosophila, absent from
RT   Saccharomyces, encodes an unusual protein with tubulin-like and
RT   myosin-like peptide motifs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:5189-5194(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTONS;
RA   MALESZKA R., DE COUET H.G., MIKLOS G.L.G.;
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1998).
DR   EMBL; AF017777; AAC28409.1; -.
DR   FLYBASE; FBgn0003464; sol.
DR   PFAM; PF00648; Peptidase_C2; 1.
DR   PFAM; PF00641; zf-RanBP; 6.
SQ   SEQUENCE   1597 AA;  174696 MW;  C6C4952E CRC32;


  Query Match             53.8%;  Score 43;  DB 5;  Length 1597;
  Best Local Similarity   61.5%;  Pred. No. 25;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SGTMRTRHSTGGT 14
          || :  ||||||:
Db    873 SGAIPKRHSTGGS 885


RESULT   4
Q65689
ID   Q65689      PRELIMINARY;      PRT;  572 AA.
AC   Q65689;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HEMAGGLUTININ-NEURAMINIDASE PROTEIN.
GN   HN.
OS   Bovine parainfluenza 3 virus.
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95287485.
RA   BREKER-KLASSEN M.M., YOO D., MITTAL S.K., SORDEN S.D., HAINES D.M.,
RA   BABIUK L.A.;
RT   "Recombinant type 5 adenoviruses expressing bovine parainfluenza
RT   virus type 3 glycoproteins protect Sigmodon hispidus cotton rats from
RT   bovine parainfluenza virus type 3 infection.";
RL   J. Virol. 69:4308-4315(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   BREKER-KLASSEN M.M.;
RL   Thesis (1995), Veterinary Infectious Disease Organization,
RL   University of Saskatchewan, Canada.
DR   EMBL; U31671; AAB03692.1; -.
DR   PFAM; PF00423; HN; 1.
SQ   SEQUENCE   572 AA;  64668 MW;  A16219EB CRC32;


  Query Match             51.2%;  Score 41;  DB 12;  Length 572;
  Best Local Similarity   61.5%;  Pred. No. 20;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 GTMRTRHSTGGTN 15
          || | |||:  ||
Db     18 GTTRDRHSSKATN 30


RESULT   5
Q16248
ID   Q16248      PRELIMINARY;      PRT;  534 AA.
AC   Q16248;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   P59FYN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95123064.
RA   RIGLEY K., SLOCOMBE P., PROUDFOOT K., WAHID S., MANDAIR K.,
RA   BEBBINGTON C.;
RT   "Human p59fyn(T) regulates OKT3-induced calcium influx by a mechanism
RT   distinct from PIP2 hydrolysis in Jurkat T cells.";
RL   J. Immunol. 154:1136-1145(1995).
DR   EMBL; S74774; AAB33113.1; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
```

```
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
SQ   SEQUENCE   534 AA;  60226 MW;  D3941652 CRC32;


  Query Match             50.0%;  Score 40;  DB 4;  Length 534;
  Best Local Similarity   61.5%;  Pred. No. 28;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SGTMRTRHSTGGT 14
          :||:|||  || |
Db     73 TGTLRTRGGTGVT 85


RESULT   6
Q25289
ID   Q25289      PRELIMINARY;      PRT;   584 AA.
AC   Q25289;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MAJOR SURFACE GLYCOPROTEIN PRECURSOR.
GN   GP63.
OS   Leishmania guyanensis.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHOM/BR/75/M4147A;
RX   MEDLINE; 94187792.
RA   STEINKRAUS H.B., GREER J.M., STEPHENSON D.C., LANGER P.J.;
RT   "Sequence heterogeneity and polymorphic gene arrangements of the "
RT   Leishmania guyanensis gp63 genes.";
RL   Mol. Biochem. Parasitol. 62:173-185(1993).
DR   EMBL; L16777; AAA29239.1; -.
DR   PFAM; PF01457; Peptidase_M8; 1.
KW   Signal.
FT   SIGNAL        1     37       POTENTIAL.
FT   CHAIN        99    577       MAJOR SURFACE GLYCOPROTEIN.
SQ   SEQUENCE   584 AA;  63896 MW;  6AA97A3E CRC32;


  Query Match             50.0%;  Score 40;  DB 5;  Length 584;
  Best Local Similarity   70.0%;  Pred. No. 30;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 GTMRTRHSTG 12
          || | :||||
Db    491 GTFRPKHSTG 500


RESULT   7
O01349
ID   O01349      PRELIMINARY;      PRT;  1921 AA.
AC   O01349;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KINESIN-73.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97188425.
RA   LI H.P., LIU Z.M., NIRENBERG M.;
RT   "Kinesin-73 in the nervous system of Drosophila embryos.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:1086-1091(1997).
DR   EMBL; U81788; AAB50404.1; -.
DR   FLYBASE; FBgn0019968; Khc-73.
DR   PFAM; PF01302; CAP_GLY; 1.
DR   PFAM; PF00225; kinesin; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS00845; CAP_GLY; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
SQ   SEQUENCE  1921 AA;  215047 MW;  BF55A112 CRC32;


  Query Match             50.0%;  Score 40;  DB 5;  Length 1921;
  Best Local Similarity   63.6%;  Pred. No. 1e+02;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 GTMRTRHSTGG 13
          | :|:||| ||
Db   1432 GVIRSRHSFGG 1442


RESULT   8
Q62844
ID   Q62844      PRELIMINARY;      PRT;   537 AA.
AC   Q62844;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROTO-ONCOGENE FYN.
GN   P59FYN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHRSP STROKE-PRONE SPONTANEOUSLY HYPERTENSIVE;
RC   TISSUE=WHOLE BRAIN;
RA   NEMOTO K., SEKIMOTO M., KAGEYAMA H., FUKAMACHI K., NEMOTO F.,
RA   UEYAMA T., SENBA E., TOMITA I.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U35365; AAA82942.1; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00017; SH2; 1.
DR   PFAM; PF00018; SH3; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
SQ   SEQUENCE   537 AA;  60701 MW;  D0A1DD46 CRC32;


  Query Match             50.0%;  Score 40;  DB 11;  Length 537;
  Best Local Similarity   61.5%;  Pred. No. 28;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 SGTMRTRHSTGGT 14
          :||:|||  || |
Db     73 TGTLRTRGGTGVT 85


RESULT   9
P90363
ID   P90363      PRELIMINARY;      PRT;    69 AA.
AC   P90363;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GLYCOPROTEIN B (FRAGMENT).
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERLIN H;
RA   BINDER T., SIEGERT W., SCHMIDT C.A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X99848; CAA68158.1; -.
FT   NON_TER       1      1
FT   NON_TER      69     69
SQ   SEQUENCE   69 AA;  7905 MW;  F68398B7 CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 69;
```

```
  Best Local Similarity   80.0%;  Pred. No. 3.5;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 RTRHSTGGTN 15
           ||| || |||
Db      25 RTRRSTDGTN 34


RESULT  10
P90292
ID   P90292       PRELIMINARY;      PRT;    69 AA.
AC   P90292;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GLYCOPROTEIN B (FRAGMENT).
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERLIN E;
RA   BINDER T., SIEGERT W., SCHMIDT C.A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X99846; CAA68156.1; -.
FT   NON_TER       1      1
FT   NON_TER      69     69
SQ   SEQUENCE   69 AA;  7905 MW;  F68398B7 CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 69;
  Best Local Similarity   80.0%;  Pred. No. 3.5;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 RTRHSTGGTN 15
           ||| || |||
Db      25 RTRRSTDGTN 34


RESULT  11
Q68570
ID   Q68570       PRELIMINARY;      PRT;    83 AA.
AC   Q68570;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCOPROTEIN B (FRAGMENT).
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IV.1;
RX   MEDLINE; 96365443.
RA   LASRY S., DENY P., ASSELOT C., RAUZY M., BOUCHER J., GUYOT C.,
RA   LEROUX M.C., LIVARTOWSKI A., REINERT P., NICOLAS J.C.;
RT   "Interstrain variations in the cytomegalovirus (CMV) glycoprotein B
RT   gene sequence among CMV-infected children attending six day care
RT   centers.";
RL   J. Infect. Dis. 174:606-609(1996).
DR   EMBL; U52134; AAB53249.1; -.
FT   NON_TER       1      1
FT   NON_TER      83     83
SQ   SEQUENCE   83 AA;  9638 MW;  72E2051D CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 83;
  Best Local Similarity   80.0%;  Pred. No. 4.3;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 RTRHSTGGTN 15
           ||| || |||
Db      29 RTRRSTDGTN 38


RESULT  12
P89538
ID   P89538       PRELIMINARY;      PRT;   171 AA.
AC   P89538;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).
GN   GP55.
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1553-91;
RX   MEDLINE; 96262011.
RA   SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,
RA   PERGOLIZZI R.;
RT   "Cytomegalovirus glycoprotein B groups associated with retinitis in
RT   AIDS.";
RL   J. Infect. Dis. 174:184-187(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1553-91;
RA   SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88701; AAB48525.1; -.
FT   NON_TER       1      1
FT   NON_TER     171    171
SQ   SEQUENCE   171 AA;  19461 MW;  CB157571 CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 171;
  Best Local Similarity   80.0%;  Pred. No. 8.8;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 RTRHSTGGTN 15
           ||| || |||
Db      97 RTRRSTDGTN 106


RESULT  13
P90392
ID   P90392       PRELIMINARY;      PRT;   170 AA.
AC   P90392;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).
GN   GP55.
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=419-91;
RX   MEDLINE; 96262011.
RA   SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,
RA   PERGOLIZZI R.;
RT   "Cytomegalovirus glycoprotein B groups associated with retinitis in
RT   AIDS.";
RL   J. Infect. Dis. 174:184-187(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=419-91;
RA   SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88700; AAB48524.1; -.
FT   NON_TER       1      1
FT   NON_TER     170    170
SQ   SEQUENCE   170 AA;  19413 MW;  FE0BE0AD CRC32;
```

```
  Query Match             50.0%;  Score 40;  DB 12;  Length 170;
  Best Local Similarity   80.0%;  Pred. No. 8.8;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 RTRHSTGGTN 15
          ||| || |||
Db     97 RTRRSTDGTN 106


RESULT  14
P90393
ID   P90393      PRELIMINARY;      PRT;   169 AA.
AC   P90393;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).
GN   GP55.
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1712-91;
RX   MEDLINE; 96262011.
RA   SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,
RA   PERGOLIZZI R.;
RT   "Cytomegalovirus glycoprotein B groups associated with retinitis in
RT   AIDS.";
RL   J. Infect. Dis. 174:184-187(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1712-91;
RA   SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88703; AAB48527.1; -.
FT   NON_TER       1       1
FT   NON_TER     169     169
SQ   SEQUENCE   169 AA;  19284 MW;  C8D31BF5 CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 169;
  Best Local Similarity   80.0%;  Pred. No. 8.7;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 RTRHSTGGTN 15
          ||| || |||
Db     96 RTRRSTDGTN 105


RESULT  15
P90394
ID   P90394      PRELIMINARY;      PRT;   172 AA.
AC   P90394;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).
GN   GP55.
OS   Human cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2515-92;
RX   MEDLINE; 96262011.
RA   SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,
RA   PERGOLIZZI R.;
RT   "Cytomegalovirus glycoprotein B groups associated with retinitis in
RT   AIDS.";
RL   J. Infect. Dis. 174:184-187(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2515-92;
RA   SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88704; AAB48528.1; -.
FT   NON_TER       1       1
FT   NON_TER     172     172
SQ   SEQUENCE   172 AA;  19597 MW;  5DC159A7 CRC32;


  Query Match             50.0%;  Score 40;  DB 12;  Length 172;
  Best Local Similarity   80.0%;  Pred. No. 8.9;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 RTRHSTGGTN 15
          ||| || |||
Db     97 RTRRSTDGTN 106


Search completed: January 12, 2000, 23:15:25
Job time: 188 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:38 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-4
Perfect score: 76
Sequence: 1 TPMFLLSRNTGEVRT 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	76	100.0	999	1	DSG3_HUMAN	P32926	homo sapien
2	64	84.2	1049	1	DSG1_HUMAN	Q02413	homo sapien
3	56	73.7	1043	1	DSG1_BOVIN	Q03763	bos taurus
4	43	56.6	1117	1	DSG2_HUMAN	Q14126	homo sapien
5	40	52.6	571	1	ASN1_YEAST	P49089	saccharomyc
6	40	52.6	571	1	ASN2_YEAST	P49090	saccharomyc
7	40	52.6	375	1	CISY_MYCSM	P26491	mycobacteri
8	38	50.0	585	1	ASN1_LOTJA	P49092	lotus japon
9	38	50.0	585	1	ASN1_PEA	P19251	pisum sativ
10	38	50.0	585	1	ASN2_LOTJA	P49093	lotus japon
11	38	50.0	582	1	ASN2_PEA	P19252	pisum sativ
12	38	50.0	583	1	ASNS_ARATH	P49078	arabidopsis
13	38	50.0	585	1	ASNS_BRAOL	P49091	brassica ol
14	38	50.0	585	1	ASNS_MAIZE	P49094	zea mays (m
15	38	50.0	590	1	ASNS_ORYSA	Q43011	oryza sativ
16	38	50.0	524	1	ASNS_SANAU	O24338	sandersonia
17	38	50.0	585	1	ASNS_TRIVS	O24661	triphysaria
18	37	48.7	186	1	ATPD_RHOBL	P05437	rhodopseudo
19	37	48.7	389	1	CISZ_ECOLI	P31660	escherichia
20	37	48.7	389	1	CISZ_SALTY	Q56063	salmonella
21	37	48.7	426	1	ENO_HELPY	P48285	helicobacte
22	37	48.7	576	1	RECN_BACSU	P17894	bacillus su
23	37	48.7	108	1	Y388_MYCPN	P75209	mycoplasma
24	36	47.4	882	1	CAD1_HUMAN	P12830	homo sapien
25	36	47.4	790	1	CAD6_HUMAN	P55285	homo sapien
26	36	47.4	789	1	CAD6_RAT	P55280	rattus norv
27	36	47.4	730	1	CADF_MOUSE	P33146	mus musculu
28	36	47.4	369	1	MURB_MYCTU	Q11148	mycobacteri
29	36	47.4	733	1	SUF_DROME	P25991	drosophila
30	36	47.4	152	1	VE6_HPV03	P36799	human papil
31	35	46.1	1314	1	ADR6_YEAST	P09547	saccharomyc
32	35	46.1	1286	1	AIDA_ECOLI	Q03155	escherichia
33	35	46.1	554	1	ASNB_ECOLI	P22106	escherichia
34	35	46.1	589	1	ASNS_ASPOF	P31752	asparagus o
35	35	46.1	379	1	CISY_ABDS2	O34002	antarctic b
36	35	46.1	396	1	CYB_PETMA	Q35534	petromyzon
37	35	46.1	488	1	DSBD_ECOLI	P36655	escherichia
38	35	46.1	5147	1	FAT_DROME	P33450	drosophila
39	35	46.1	609	1	HAPT_VIBCH	P24153	vibrio chol
40	35	46.1	1538	1	LHR_ECOLI	P30015	escherichia
41	35	46.1	219	1	NODB_BRASP	P04675	bradyrhizob
42	35	46.1	1179	1	RPOB_MYCLE	P30760	mycobacteri
43	35	46.1	1178	1	RPOB_MYCTU	P47766	mycobacteri
44	35	46.1	665	1	Y176_METTH	O26278	methanobact
45	35	46.1	851	1	YD72_SCHPO	Q10327	schizosacch

ALIGNMENTS

RESULT 1
DSG3_HUMAN
ID DSG3_HUMAN STANDARD; PRT; 999 AA.
AC P32926;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92069753.
RA AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion.";
RL Cell 67:869-877(1991).
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC ----------

CC ----------
DR EMBL; M76482; AAA60230.1; -.
DR PIR; A41088; IJHUG3.
DR HSSP; P09803; 1EDH.
DR MIM; 169615; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.

FT	SIGNAL	1	23 POTENTIAL.
FT	PROPEP	24	49 POTENTIAL.
FT	CHAIN	50	999 DESMOGLEIN 3.
FT	DOMAIN	50	615 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	616	640 POTENTIAL.
FT	DOMAIN	641	999 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50	158 CADHERIN 1.
FT	REPEAT	159	268 CADHERIN 2.
FT	REPEAT	269	383 CADHERIN 3.
FT	REPEAT	386	499 CADHERIN 4.
FT	REPEAT	910	935 DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966 DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110	110 POTENTIAL.

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match             100.0%;  Score 76;  DB 1;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEVRT 15
          |||||||||||||||
Db    206 TPMFLLSRNTGEVRT 220


RESULT   2
DSG1_HUMAN
ID   DSG1_HUMAN     STANDARD;      PRT;  1049 AA.
AC   Q02413;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN   DSG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 91271279.
RA   WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOTIS P., POYNTER D.,
RA   ARNEMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA   BUXTON R.S., MAGEE A.I.;
RT   "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT   junctions, is related to the cadherin family of cell adhesion
RT   molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56654; CAA39976.1; -.
DR   PIR; S16906; IJHUG1.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125670; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1049       DESMOGLEIN 1.
FT   DOMAIN       50    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    570       POTENTIAL.
FT   DOMAIN      571   1049       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    497       CADHERIN 4.
FT   REPEAT      813    839       DESMOGLEIN REPEAT 1.
FT   REPEAT      840    869       DESMOGLEIN REPEAT 2.
FT   REPEAT      870    899       DESMOGLEIN REPEAT 3.
FT   REPEAT      900    927       DESMOGLEIN REPEAT 4.
FT   REPEAT      928    956       DESMOGLEIN REPEAT 5.
FT   DOMAIN      969   1019       GLY/SER-RICH.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    180    180       POTENTIAL.
SQ   SEQUENCE   1049 AA;  113715 MW;  FDD79961 CRC32;


  Query Match             84.2%;  Score 64;  DB 1;  Length 1049;
  Best Local Similarity   66.7%;  Pred. No. 0.00061;
  Matches   10;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEVRT 15
          :|||:::||||||:|
Db    206 SPMFIINRNTGEIRT 220


RESULT   3
DSG1_BOVIN
ID   DSG1_BOVIN     STANDARD;      PRT;  1043 AA.
AC   Q03763;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
GN   DSG1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANKE W.W.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 44-1043 FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 91168965.
RA   KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA   ZIMBELMANN R., FRANKE W.W.;
RT   "Identification of desmoglein, a constitutive desmosomal
RT   glycoprotein, as a member of the cadherin family of cell adhesion
RT   molecules.";
RL   Eur. J. Cell Biol. 53:1-12(1990).
RN   [3]
RP   REVISIONS, AND SEQUENCE OF 101-123.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA   FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
RN   [4]
RP   SEQUENCE OF 44-493 FROM N.A.
RX   MEDLINE; 91097553.
RA   GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT   "Desmoglein shows extensive homology to the cadherin family of cell
RT   adhesion molecules.";
RL   Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X58466; CAA41380.1; -.
DR   EMBL; X57784; CAA40930.1; -.
DR   EMBL; M58165; AAA62709.1; -.
DR   PIR; S14603; IJBOG1.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24     49       POTENTIAL.
FT   CHAIN       50   1043       DESMOGLEIN 1.
FT   DOMAIN      50    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   549    573       POTENTIAL.
FT   DOMAIN     574   1043       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      50    158       CADHERIN 1.
FT   REPEAT     159    270       CADHERIN 2.
FT   REPEAT     271    385       CADHERIN 3.
FT   REPEAT     386    498       CADHERIN 4.
FT   REPEAT     819    845       DESMOGLEIN REPEAT 1.
FT   REPEAT     846    875       DESMOGLEIN REPEAT 2.
FT   REPEAT     876    905       DESMOGLEIN REPEAT 3.
FT   REPEAT     906    933       DESMOGLEIN REPEAT 4.
FT   REPEAT     934    962       DESMOGLEIN REPEAT 5.
FT   DOMAIN     963   1012       GLY/SER-RICH.
FT   CARBOHYD   110    110
FT   CARBOHYD   180    180       POTENTIAL.
FT   CARBOHYD   496    496       POTENTIAL.
FT   CONFLICT   124    124       I -> V (IN REF. 4).
SQ   SEQUENCE  1043 AA;  112243 MW;  13898584 CRC32;


  Query Match             73.7%;  Score 56;  DB 1;  Length 1043;
  Best Local Similarity   60.0%;  Pred. No. 0.018;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSRNTGEVRT 15
           :|||:::| |||:||
Db     206 SPMFIINRYTGEIRT 220


RESULT   4
DSG2_HUMAN
ID   DSG2_HUMAN     STANDARD;      PRT;  1117 AA.
AC   Q14126;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 2 PRECURSOR (HDGC).
GN   DSG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CARCINOMA;
RX   MEDLINE; 94192736.
RA   SCHAEFER S., KOCH P.J., FRANKE W.W.;
RT   "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT   expression catalogue of the desmoglein subfamily of desmosomal
RT   cadherins.";
RL   Exp. Cell Res. 211:391-399(1994).
RN   [2]
RP   SEQUENCE OF 777-1117 FROM N.A.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z26317; CAA81226.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 125671; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24     48       POTENTIAL.
FT   CHAIN       49   1117       DESMOGLEIN 2.
FT   DOMAIN      49    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    633       POTENTIAL.
FT   DOMAIN     634   1117       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      49    159       CADHERIN 1.
FT   REPEAT     160    272       CADHERIN 2.
FT   REPEAT     273    387       CADHERIN 3.
FT   REPEAT     388    502       CADHERIN 4.
FT   REPEAT     880    911       DESMOGLEIN REPEAT 1.
FT   REPEAT     912    941       DESMOGLEIN REPEAT 2.
FT   REPEAT     942    967       DESMOGLEIN REPEAT 3.
FT   REPEAT     968    991       DESMOGLEIN REPEAT 4.
FT   REPEAT     992   1020       DESMOGLEIN REPEAT 5.
FT   REPEAT    1021   1050       DESMOGLEIN REPEAT 6.
FT   CARBOHYD   111    111       POTENTIAL.
FT   CARBOHYD   181    181       POTENTIAL.
FT   CARBOHYD   308    308       POTENTIAL.
FT   CARBOHYD   461    461       POTENTIAL.
FT   CARBOHYD   513    513       POTENTIAL.
SQ   SEQUENCE  1117 AA;  122385 MW;  84D3B898 CRC32;


  Query Match             56.6%;  Score 43;  DB 1;  Length 1117;
  Best Local Similarity   50.0%;  Pred. No. 4.8;
  Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEVRT 15
           |:| |:::|||: |
Db     208 PVFYLNKDTGEIYT 221


RESULT   5
ASN1_YEAST
ID   ASN1_YEAST     STANDARD;      PRT;   571 AA.
AC   P49089;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 1 (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).
GN   ASN1 OR YPR145W OR P9659.3.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
```

```
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DANG V.D., BOLOTIN-FUKUHARA M., DAIGNAN-FORNIER B.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA   FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
RA   HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
RA   JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
RA   MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
RA   TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,
RA   WILSON R., WATERSTON R.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48675; CAA88594.1; -.
DR   EMBL; U40829; AAB68284.1; -.
DR   SGD; L0002732; ASN1.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   571 AA;  64339 MW;  6FA2535B CRC32;


  Query Match             52.6%;  Score 40;  DB 1;  Length 571;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           ||||||||
Db     346 TPMFLLSR 353


RESULT   6
ASN2_YEAST
ID   ASN2_YEAST     STANDARD;      PRT;   571 AA.
AC   P49090;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
GN   ASN2 OR YGR124W OR G6358.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 97197982.
RA   VAN DYCK L., TETTELIN H., PURNELLE B., GOFFEAU A.;
RT   "An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT   unknown open reading frames, the gene for an Asn synthase, remnants
RT   of Ty and three tRNA genes.";
RL   Yeast 13:171-176(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   VAN DYCK L., SKALA J., DE WERGIFOSSE P., PURNELLE B., TALLA E.,
RA   NAWROCKI A., DEL BINO S., GOFFEAU A.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83099; CAA58159.1; -.
DR   EMBL; Z72909; CAA97135.1; -.
DR   SGD; L0003156; ASN2.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   571 AA;  64461 MW;  926C9736 CRC32;


  Query Match             52.6%;  Score 40;  DB 1;  Length 571;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           ||||||||
Db     345 TPMFLLSR 352


RESULT   7
CISY_MYCSM
ID   CISY_MYCSM     STANDARD;      PRT;   375 AA.
AC   P26491;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   CITRATE SYNTHASE (EC 4.1.3.7).
GN   GLTA.
OS   Mycobacterium smegmatis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 607;
RX   MEDLINE; 91354207.
RA   DAVID M., LUBINSKY-MINK S., BENZVI A., SUISSA M., UNITZUR S.,
RA   KUHN J.C.;
RT   "Citrate synthase from Mycobacterium smegmatis. Cloning, sequence
RT   determination and expression in Escherichia coli.";
RL   Biochem. J. 278:225-234(1991).
CC   -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC       OXALOACETATE.
CC   -!- ENZYME REGULATION: ALLOSTERICALLY INHIBITED BY NADH.
CC   -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC   -!- SUBUNIT: HOMOHEXAMER.
CC   -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC       CAPABLE OF OXIDATIVE METABOLISM.
CC   -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X60513; CAA43028.1; -.
DR   PIR; S17168; YKMY.
DR   HSSP; Q53554; 1AJ8.
DR   PFAM; PF00285; citrate_synt; 1.
DR   PROSITE; PS00480; CITRATE_SYNTHASE; 1.
KW   Lyase; Tricarboxylic acid cycle; Allosteric enzyme.
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   ACT_SITE    317    317       BY SIMILARITY.
SQ   SEQUENCE   375 AA;  41501 MW;  6121D32A CRC32;


  Query Match             52.6%;  Score 40;  DB 1;  Length 375;
  Best Local Similarity   63.6%;  Pred. No. 5.4;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTG 11
          ||:|:|| |
Db    335 TPLFVMSRITG 345


RESULT   8
ASN1_LOTJA
ID   ASN1_LOTJA     STANDARD;      PRT;   585 AA.
AC   P49092;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 1 (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).
GN   AS1.
OS   Lotus japonicus.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Lotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. GIFU / B-129;
RX   MEDLINE; 96270368.
RA   WATERHOUSE R.N., SMYTH A.J., MASSONEAU A., PROSSER I.M.,
RA   CLARKSON D.T.;
RT   "Molecular cloning and characterisation of asparagine synthetase from
RT   Lotus japonicus: dynamics of asparagine synthesis in N-sufficient
RT   conditions.";
RL   Plant Mol. Biol. 30:883-897(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X89409; CAA61589.1; -.
DR   HSSP; P17169; 1GMS.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   585 AA;  66330 MW;  7FDC0436 CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          |||||:||
Db    321 TPMFLMSR 328


RESULT   9
ASN1_PEA
ID   ASN1_PEA       STANDARD;      PRT;   585 AA.
AC   P19251; O49925;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ASPARAGINE SYNTHETASE, NODULE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
GN   AS1.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Pisum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SPARKLE; TISSUE=ROOT NODULES;
RX   MEDLINE; 90151604.
RA   TSAI F.Y., CORUZZI G.M.;
RT   "Dark-induced and organ-specific expression of two asparagine
RT   synthetase genes in Pisum sativum.";
RL   EMBO J. 9:323-332(1990).
RN   [2]
RP   SEQUENCE OF 1-83 FROM N.A.
RC   STRAIN=CV. FELTHAM FIRST;
RA   NGAI N., TSAI F.Y., CORUZZI G.M.;
RT   "Light-induced transcriptional repression of the pea AS1 gene:
RT   identification of cis-elements and transfactors.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- TISSUE SPECIFICITY: NODULE.
CC   -!- INDUCTION: DARK-INDUCED.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X52179; CAA36429.1; -.
DR   EMBL; Y13321; CAA73762.1; -.
DR   PIR; S11444; AJPMN1.
DR   HSSP; P17169; 1GMS.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
```

```
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   585 AA;  66222 MW;  9AB4FBC5 CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          ||||||:||
Db    322 TPMFLMSR 329


RESULT  10
ASN2_LOTJA
ID   ASN2_LOTJA     STANDARD;      PRT;   585 AA.
AC   P49093;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
GN   AS2.
OS   Lotus japonicus.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Lotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. GIFU / B-129;
RX   MEDLINE; 96270368.
RA   WATERHOUSE R.N., SMYTH A.J., MASSONEAU A., PROSSER I.M.,
RA   CLARKSON D.T.;
RT   "Molecular cloning and characterisation of asparagine synthetase from
RT   Lotus japonicus: dynamics of asparagine synthesis in N-sufficient
RT   conditions.";
RL   Plant Mol. Biol. 30:883-897(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89410; CAA61590.1; -.
DR   HSSP; P17169; 1GMS.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   585 AA;  65838 MW;  2FE40574 CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          ||||||:||
Db    321 TPMFLMSR 328


RESULT  11
ASN2_PEA
ID   ASN2_PEA       STANDARD;      PRT;   582 AA.
AC   P19252; O49926;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ASPARAGINE SYNTHETASE, ROOT [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
DE   (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
GN   AS2.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Pisum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SPARKLE; TISSUE=ROOT;
RX   MEDLINE; 90151604.
RA   TSAI F.Y., CORUZZI G.M.;
RT   "Dark-induced and organ-specific expression of two asparagine
RT   synthetase genes in Pisum sativum.";
RL   EMBO J. 9:323-332(1990).
RN   [2]
RP   SEQUENCE OF 1-83 FROM N.A.
RC   STRAIN=CV. FELTHAM FIRST;
RA   NGAI N., TSAI F.Y., CORUZZI G.M.;
RT   "Light-induced transcriptional repression of the pea AS1 gene:
RT   identification of cis-elements and transfactors.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- TISSUE SPECIFICITY: ROOTS.
CC   -!- INDUCTION: DARK-INDUCED.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52180; CAA36430.1; -.
DR   EMBL; Y13322; CAA73763.1; -.
DR   PIR; S11443; AJPMN2.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
FT   CONFLICT     33     33       E -> D (IN REF. 2).
FT   CONFLICT     44     44       Y -> F (IN REF. 2).
FT   CONFLICT     47     47       Q -> H (IN REF. 2).
FT   CONFLICT     77     77       I -> L (IN REF. 2).
SQ   SEQUENCE   582 AA;  65518 MW;  793421FA CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 582;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          ||||||:||
Db    321 TPMFLMSR 328
```

```
RESULT  12
ASNS_ARATH
ID   ASNS_ARATH     STANDARD;      PRT;   583 AA.
AC   P49078;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE   DEPENDENT ASPARAGINE SYNTHETASE).
GN   ASN1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE; 95148732.
RA   LAM H.M., PENG S.S., CORUZZI G.M.;
RT   "Metabolic regulation of the gene encoding glutamine-dependent ...
RT   asparagine synthetase in Arabidopsis thaliana.";
RL   Plant Physiol. 106:1347-1357(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29083; AAA74359.1; -.
DR   HSSP; P17169; 1GMS.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW   Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   583 AA;  65489 MW;  F8FC9672 CRC32;


  Query Match            50.0%;  Score 38;  DB 1;  Length 583;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     321 TPMFLMSR 328


RESULT  13
ASNS_BRAOL
ID   ASNS_BRAOL     STANDARD;      PRT;   585 AA.
AC   P49091;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE   DEPENDENT ASPARAGINE SYNTHETASE).
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Brassica.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SHOGUN;
RA   DOWNS C.G., POGSON B.J., DAVIES K.M., ALMIRA E.C.;
RT   "An asparagine synthetase cDNA clone from Broccoli (Brassica oleracea
RT   L.).";
RL   (In) Plant Gene Register PGR95-016.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X84448; CAA59138.1; -.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   585 AA;  65541 MW;  B6DCFB50 CRC32;


  Query Match            50.0%;  Score 38;  DB 1;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     322 TPMFLMSR 329


RESULT  14
ASNS_MAIZE
ID   ASNS_MAIZE     STANDARD;      PRT;   585 AA.
AC   P49094;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE   DEPENDENT ASPARAGINE SYNTHETASE).
GN   ASN1 OR AS.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Zea.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DEA; TISSUE=ROOT MERISTEM;
RX   MEDLINE; 96158342.
RA   CHEVALIER C., BOURGEOIS E., JUST D., RAYMOND P.;
RT   "Metabolic regulation of asparagine synthetase gene expression in
RT   maize (Zea mays L.) root tips.";
RL   Plant J. 9:1-11(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X82849; CAA58052.1; -.
DR   MAIZEDB; 79071; -.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   585 AA;  66446 MW;  B24B0478 CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
             |||||:||
Db     321 TPMFLMSR 328


RESULT  15
ASNS_ORYSA
ID   ASNS_ORYSA     STANDARD;      PRT;   590 AA.
AC   Q43011;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE   DEPENDENT ASPARAGINE SYNTHETASE).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE; TISSUE=CALLUS;
RA   SUEYOSHI K., KAWACHI T., NAKAJIMA A., YAMAGATA H., SUGIMOTO T.,
RA   IWASAKI T., OJI Y.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC   -!- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U55873; AAB03991.1; -.
DR   PFAM; PF00310; GATase_2; 1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Ligase; Asparagine biosynthesis; Glutamine amidotransferase.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
SQ   SEQUENCE   590 AA;  66096 MW;  33CDFF30 CRC32;


  Query Match             50.0%;  Score 38;  DB 1;  Length 590;
  Best Local Similarity   87.5%;  Pred. No. 21;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
             |||||:||
Db     321 TPMFLMSR 328


Search completed: January  7, 2000, 13:25:39
Job time: 391 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:41 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-4
Perfect score: 76
Sequence: 1 TPMFLLSRNTGEVRT 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	76	100.0	999	1	IJHUG3	desmoglein 3 precu
2	64	84.2	1049	1	IJHUG1	desmoglein 1 precu
3	56	73.7	1043	1	IJBOG1	desmoglein 1 precu
4	43.5	57.2	295	2	S32892	hypothetical prote
5	43	56.6	1117	2	S38673	desmoglein 2 - hum
6	40	52.6	375	1	YKMY	citrate (si)-synth
7	40	52.6	572	2	S52694	asparagine synthas
8	40	52.6	572	2	S55982	asparagine synthas
9	40	52.6	452	2	F69294	DNA repair protein
10	39	51.3	393	2	D70539	probable citrate s
11	39	51.3	163	2	A71536	hypothetical prote
12	39	51.3	302	2	C72756	probable multiple
13	38	50.0	583	1	AJPMN2	asparagine synthas
14	38	50.0	586	1	AJPMN1	asparagine synthas
15	38	50.0	586	2	S69182	asparagine synthas
16	38	50.0	586	2	S69183	asparagine synthas
17	38	50.0	586	2	S52387	asparagine synthas
18	38	50.0	591	2	T03602	probable asparagin
19	38	50.0	586	2	T02978	asparagine synthas
20	38	50.0	581	2	T08846	asparagine synthas
21	38	50.0	584	2	T12989	asparagine synthas
22	37	48.7	576	1	B35128	DNA repair and gen
23	37	48.7	186	2	S04671	H+-transporting AT
24	37	48.7	389	2	E64760	citrate (si)-synth
25	37	48.7	426	2	S58684	phosphopyruvate hy
26	37	48.7	426	2	H71967	enolase - Helicoba
27	37	48.7	108	2	S73599	MG388 homolog D02_
28	37	48.7	367	2	T02000	hypothetical prote
29	37	48.7	852	2	B72685	hypothetical prote
30	37	48.7	365	2	E72475	hypothetical prote
31	37	48.7	211	2	E71334	hypothetical prote
32	37	48.7	593	2	T10301	ribonucleotide red
33	36	47.4	882	1	IJHUCE	cadherin 1 precurs
34	36	47.4	730	1	IJMSCM	M-cadherin - mouse
35	36	47.4	369	2	E70743	probable murB prot
36	36	47.4	790	2	I37016	cadherin-6 - human
37	36	47.4	789	2	I52701	K-cadherin - rat
38	36	47.4	790	2	I50178	cadherin-6B - chic
39	36	47.4	483	2	T06459	62K sucrose-bindin
40	36	47.4	152	2	S36550	E6 protein - human
41	36	47.4	379	2	I49020	retinoid X recepto
42	36	47.4	484	2	I49018	retinoid X recepto
43	36	47.4	469	2	A56918	farnesoid x-activa
44	36	47.4	466	2	H64904	hypothetical prote
45	36	47.4	1417	2	T00661	hypothetical prote

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 76; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 7.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTGEVRT 15
||||||||||||||||
Db 206 TPMFLLSRNTGEVRT 220

RESULT 2
IJHUG1
desmoglein 1 precursor - human
N;Alternate names: desmosomal glycoprotein I
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S16906; A39706; A61254; A61279; S16158
R;Buxton, R.S.
submitted to the EMBL Data Library, November 1990
A;Reference number: S16906
A;Accession: S16906
A;Molecule type: mRNA

A;Residues: 1-1049 <BUX>
A;Cross-references: EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
R;Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; R
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991
A;Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,
A;Reference number: A39706; MUID:91271279
A;Accession: A39706
A;Molecule type: mRNA
A;Residues: 24-1049 <WHE>
A;Cross-references: GB:X56654
R;Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
J. Cell Sci. 99, 809-821, 1991
A;Title: Structural analysis and expression of human desmoglein: a cadherin-like compone
A;Reference number: A61254; MUID:92121251
A;Accession: A61254
A;Molecule type: mRNA
A;Residues: 26-1049 <NIL>
R;Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee
Biochem. Soc. Trans. 19, 1060-1064, 1991
A;Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfami
A;Reference number: A61279; MUID:92175187
A;Accession: A61279
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-55 <WH3>
C;Genetics:
A;Gene: GDB:DSG1
A;Cross-references: GDB:126563; OMIM:125670
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1049/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-493/Domain: cadherin repeat homology <CR4>
F;509-530/Region: serine/threonine-rich
F;549-569/Domain: transmembrane #status predicted <TMM>
F;572-1049/Domain: intracellular #status predicted <INT>
F;840-869/Domain: desmoglein repeat <DG1>
F;870-899/Domain: desmoglein repeat <DG2>
F;900-927/Domain: desmoglein repeat <DG3>
F;928-956/Domain: desmoglein repeat <DG4>
F;969-1019/Region: glycine/serine-rich
F;110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.2%; Score 64; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 0.0013;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TPMFLLSRNTGEVRT 15
           :|||:::|||||:||
Db     206 SPMFIINRNTGEIRT 220
```

RESULT 3
IJBOG1
desmoglein 1 precursor - bovine
N;Alternate names: desmoglein BDGM
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S14603; A38872; A37785; S38721; A48173; S24412
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
submitted to the EMBL Data Library, March 1991
A;Description: Complete sequence of the desmoglein precursor and evidence for the existe
A;Reference number: S14603
A;Accession: S14603
A;Molecule type: mRNA
A;Residues: 1-1043 <KOC>
A;Cross-references: EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti
A;Reference number: A38872; MUID:92037656
A;Accession: A38872
A;Molecule type: mRNA
A;Residues: 1-87;968-1043 <KO2>
A;Cross-references: GB:S64268; GB:S64270
R;Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990
A;Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion
A;Reference number: A37785; MUID:91097553
A;Accession: A37785
A;Molecule type: mRNA
A;Residues: 44-123,'V',125-493 <GOO>
A;Cross-references: GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
R;Zimbelmann, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S38721
A;Accession: S38721
A;Molecule type: mRNA
A;Residues: 44-1043 <ZIM>
A;Cross-references: EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
R;Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.
Eur. J. Cell Biol. 53, 1-12, 1990
A;Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m
A;Reference number: A48173; MUID:91168965
A;Accession: A48173
A;Molecule type: mRNA
A;Residues: 44-1001,'AQPPSAT' <KO3>
A;Cross-references: GB:X57784
A;Note: this sequence has been revised in references A38872 and S38721
C;Genetics:
A;Gene: DSG1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1043/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-491/Domain: cadherin repeat homology <CR4>
F;549-574/Domain: transmembrane #status predicted <TMM>
F;575-1043/Domain: intracellular #status predicted <INT>
F;846-875/Domain: desmoglein repeat <DG1>
F;876-905/Domain: desmoglein repeat <DG2>
F;906-933/Domain: desmoglein repeat <DG3>
F;934-962/Domain: desmoglein repeat <DG4>
F;963-1012/Region: glycine/serine-rich
F;110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 73.7%; Score 56; DB 1; Length 1043;
Best Local Similarity 60.0%; Pred. No. 0.037;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 TPMFLLSRNTGEVRT 15
           :|||:::| |||:||
Db     206 SPMFIINRYTGEIRT 220
```

RESULT 4
S32892
hypothetical protein 6 (pefI 5' region) - Salmonella typhimurium
C;Species: Salmonella typhimurium
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Sep-1999
C;Accession: S32892
R;Friedrich, M.J.; Kinsey, N.E.; Vila, J.; Kadner, R.J.
Mol. Microbiol. 8, 543-558, 1993

A;Title: Nucleotide sequence of a 13.9kb segment of the 90kb virulence plasmid of Salmon
A;Reference number: S32886; MUID:93316852
A;Accession: S32892
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-295 <FRI>
A;Cross-references: EMBL:L08613
C;Superfamily: Salmonella typhimurium hypothetical protein 6 (pefI 5' region)

Query Match 57.2%; Score 43.5; DB 2; Length 295;
Best Local Similarity 80.0%; Pred. No. 1.9;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 TPMFLLSRNTGEVRT 15
| || ||| ||||||
Db 132 TLMFSLSR-TGEVRT 145

RESULT 5
S38673
desmoglein 2 - human
N;Alternate names: desmoglein HDGC
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C;Accession: S38673; B38872
R;Zimbelmann, R.
submitted to the EMBL Data Library, September 1993
A;Reference number: S38673
A;Accession: S38673
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1117 <ZIM>
A;Cross-references: EMBL:Z26317; NID:g416177; PIDN:CAA81226.1; PID:g416178
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide
A;Reference number: A38872; MUID:92037656
A;Accession: B38872
A;Molecule type: mRNA
A;Residues: 777-1117 <KOC>
A;Cross-references: GB:S64273
C;Genetics:
A;Gene: GDB:DSG2
A;Cross-references: GDB:128808; OMIM:125671
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein
F;51-158/Domain: cadherin repeat homology <CR1>
F;161-271/Domain: cadherin repeat homology <CR2>

Query Match 56.6%; Score 43; DB 2; Length 1117;
Best Local Similarity 50.0%; Pred. No. 9.3;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PMFLLSRNTGEVRT 15
|:| |:::|||: |
Db 208 PVFYLNKDTGEIYT 221

RESULT 6
YKMY
citrate (si)-synthase (EC 4.1.3.7) - Mycobacterium smegmatis
C;Species: Mycobacterium smegmatis
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
C;Accession: S17168
R;David, M.; Lubinsky-Mink, S.; Ben-Zvi, A.; Suissa, M.; Ulitzur, S.; Kuhn, J.
Biochem. J. 278, 225-234, 1991
A;Title: Citrate synthase from Mycobacterium smegmatis. Cloning, sequence determination
A;Reference number: S17168; MUID:91354207
A;Accession: S17168
A;Molecule type: DNA
A;Residues: 1-375 <DAV>
A;Cross-references: EMBL:X60513; NID:g44513; PIDN:CAA43028.1; PID:g44514
C;Genetics:
A;Gene: gltA
C;Superfamily: citrate (si)-synthase
C;Keywords: acetyl-CoA; carbon-carbon lyase; oxo-acid-lyase; tricarboxylic acid cycle
F;227,266,317/Active site: His, His, Asp #status predicted

Query Match 52.6%; Score 40; DB 1; Length 375;
Best Local Similarity 63.6%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTG 11
||:|T:|T T|
Db 335 TPLFVMSRITG 345

RESULT 7
S52694
asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN1 - yeast (Saccharomyces
N;Alternate names: protein P9659.3; protein YPR145w
C;Species: Saccharomyces cerevisiae
C;Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 16-Jul-1999
C;Accession: S52694; S69033
R;Dang, V.D.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.
submitted to the EMBL Data Library, March 1995
A;Description: Multiple regulatory systems control expression of the Saccharomyces ce
A;Reference number: S52694
A;Accession: S52694
A;Molecule type: DNA
A;Residues: 1-572 <DAN>
A;Cross-references: EMBL:Z48675; NID:g1163061; PIDN:CAA88594.1; PID:g747902
R;Fulton, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of S. cerevisiae cosmid 9659.
A;Reference number: S69019
A;Accession: S69033
A;Molecule type: DNA
A;Residues: 1-572 <FUL>
A;Cross-references: EMBL:U40829; NID:g1066476; PIDN:AAB68284.1; PID:g1066479; MIPS:YP
C;Genetics:
A;Gene: SGD:ASN1
A;Cross-references: SGD:S0006349; MIPS:YPR145w
A;Map position: 16R
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)
C;Keywords: asparagine biosynthesis; ligase
F;2/Active site: Cys (amide transfer) #status predicted

Query Match 52.6%; Score 40; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSR 8
||||||||
Db 347 TPMFLLSR 354

RESULT 8
S55982
asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN2 - yeast (Saccharomyces
N;Alternate names: protein G6358; protein YGR124w
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
C;Accession: S55982; S64433
R;van Dyck, L.; Goffeau, A.
submitted to the EMBL Data Library, December 1994
A;Description: Genes for an asn synthase, a GLFG-motif nucleoporin and a putative hom
e new ORFs, remnants of Ty and three tRNA genes.
A;Reference number: S55976
A;Accession: S55982
A;Molecule type: DNA

A;Residues: 1-572 <VAN>
A;Cross-references: EMBL:X83099; NID:g642340; PIDN:CAA58159.1; PID:g642347
R;Van Dyck, L.; Skala, J.; de Wergifosse, P.; Purnelle, B.; Talla, E.; Nawrocki, A.; Del
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64428
A;Accession: S64433
A;Molecule type: DNA
A;Residues: 1-572 <VAW>
A;Cross-references: EMBL:Z72909; NID:g1323202; PIDN:CAA97135.1; PID:e243463; PID:g132320
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:ASN2
A;Cross-references: SGD:S0003356; MIPS:YGR124w
A;Map position: 7R
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)
C;Keywords: asparagine biosynthesis; ligase
F;2/Active site: Cys (amide transfer) #status predicted

Query Match 52.6%; Score 40; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSR 8
||||||||
Db 346 TPMFLLSR 353

RESULT 9
F69294
DNA repair protein RAD25 homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Sep-1999
C;Accession: F69294
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaec
A;Reference number: A69250; MUID:98049343
A;Accession: F69294
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-452 <KLE>
A;Cross-references: GB:AE001080; GB:AE000782; NID:g2689403; PIDN:AAB90879.1; PID:g265027
C;Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
C;Keywords: ATP; P-loop
F;95-417/Domain: DEAD/H box helicase homology <DEAD>
F;95-102/Region: nucleotide-binding motif A (P-loop)
F;175-180/Region: nucleotide-binding motif B
F;179-182/Region: DEXH motif

Query Match 52.6%; Score 40; DB 2; Length 452;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 5 LLSRNTGEVRT 15
|:|| |||| |
Db 430 LISRGTGEVNT 440

RESULT 10
D70539
probable citrate synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Jun-1999
C;Accession: D70539
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: D70539
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <COL>
A;Cross-references: GB:Z95585; GB:AL123456; NID:g3261787; PIDN:CAB09042.1; PID:e31715
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: gltA1
C;Superfamily: citrate (si)-synthase

Query Match 51.3%; Score 39; DB 2; Length 393;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTG 11
||:|::|| ||
Db 350 TPIFVMSRITG 360

RESULT 11
A71536
hypothetical protein CT260 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
C;Accession: A71536
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809
A;Accession: A71536
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-163 <ARN>
A;Cross-references: GB:AE001299; GB:AE001273; NID:g3328671; PID:g3328672
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT260

Query Match 51.3%; Score 39; DB 2; Length 163;
Best Local Similarity 58.3%; Pred. No. 6.6;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PMFLLSRNTGEV 13
|:| : || |||
Db 43 PLFFVIRNEGEV 54

RESULT 12
C72756
probable multiple sugar-binding transport system permease protein APE0040 - Aeropyrum
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: C72756
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: C72756
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <KAW>
A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA78949.1; PID:d1042725; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0040

```
  Query Match           51.3%;  Score 39;  DB 2;  Length 302;
  Best Local Similarity  50.0%;  Pred. No. 13;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEVRT 15
           |:||| :| | : |
Db     150 PLFLLLKNLGLINT 163


RESULT  13
AJPMN2
asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 - garden pea
N;Alternate names: asparagine synthetase (glutamine-hydrolyzing)
C;Species: Pisum sativum (garden pea)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Jun-1999
C;Accession: S11443
R;Tsai, F.Y.; Coruzzi, G.M.
EMBO J. 9, 323-332, 1990
A;Title: Dark-induced and organ-specific expression of two asparagine synthetase genes i
A;Reference number: S11443; MUID:90151604
A;Accession: S11443
A;Molecule type: DNA
A;Residues: 1-583 <TSA>
A;Cross-references: EMBL:X52180; NID:g20651; PIDN:CAA36430.1; PID:g20652
C;Genetics:
A;Gene: AS2
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)
C;Keywords: asparagine biosynthesis; ligase
F;2/Active site: Cys (amide transfer) #status predicted


  Query Match           50.0%;  Score 38;  DB 1;  Length 583;
  Best Local Similarity  87.5%;  Pred. No. 39;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     322 TPMFLMSR 329


RESULT  14
AJPMN1
asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 - garden pea
N;Alternate names: asparagine synthetase (glutamine-hydrolyzing)
C;Species: Pisum sativum (garden pea)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Jun-1999
C;Accession: S11444
R;Tsai, F.Y.; Coruzzi, G.M.
EMBO J. 9, 323-332, 1990
A;Title: Dark-induced and organ-specific expression of two asparagine synthetase genes i
A;Reference number: S11443; MUID:90151604
A;Accession: S11444
A;Molecule type: DNA
A;Residues: 1-586 <TSA>
A;Cross-references: EMBL:X52179; NID:g20649; PIDN:CAA36429.1; PID:g20650
C;Comment: This protein is one of a family of glutamine amidotransferases that have dual
: an aminator domain, which catalyzes the ammonia-dependent reaction, and a glutamine am
C;Genetics:
A;Gene: AS1
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)
C;Keywords: asparagine biosynthesis; ligase
F;175-586/Domain: aminator <AMN>
F;2/Active site: Cys (amide transfer) #status predicted


  Query Match           50.0%;  Score 38;  DB 1;  Length 586;
  Best Local Similarity  87.5%;  Pred. No. 39;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     323 TPMFLMSR 330


RESULT  15
S69182
asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 - Lotus japonicus
C;Species: Lotus japonicus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
C;Accession: S69182; S57931
R;Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.; Clarkson, D.T.
Plant Mol. Biol. 30, 883-897, 1996
A;Title: Molecular cloning and characterisation of asparagine synthetase from Lotus j
A;Reference number: S69182; MUID:96270368
A;Accession: S69182
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-586 <WAT>
A;Cross-references: EMBL:X89409; NID:g897770; PIDN:CAA61589.1; PID:g897771
A;Experimental source: strain B-129
C;Genetics:
A;Gene: AS1
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)
C;Keywords: asparagine biosynthesis; ligase


  Query Match           50.0%;  Score 38;  DB 2;  Length 586;
  Best Local Similarity  87.5%;  Pred. No. 39;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     322 TPMFLMSR 329


Search completed: January  7, 2000, 08:52:42
Job time: 2272 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:35 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-2
Perfect score: 78
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	100.0	999	1	DSG3_HUMAN	P32926 homo sapien
2	56	71.8	1117	1	DSG2_HUMAN	Q14126 homo sapien
3	55	70.5	1043	1	DSG1_BOVIN	Q03763 bos taurus
4	55	70.5	1049	1	DSG1_HUMAN	Q02413 homo sapien
5	45	57.7	893	1	DSC1_BOVIN	Q01107 bos taurus
6	45	57.7	894	1	DSC1_HUMAN	Q08554 homo sapien
7	43	55.1	782	1	CAD5_PIG	O02840 sus scrofa
8	43	55.1	790	1	CAD6_HUMAN	P55285 homo sapien
9	43	55.1	789	1	CAD6_RAT	P55280 rattus norv
10	43	55.1	789	1	CADA_CHICK	P79995 gallus gall
11	43	55.1	790	1	CADE_HUMAN	Q13634 homo sapien
12	43	55.1	566	1	YGA3_YEAST	P53197 saccharomyc
13	42	53.8	713	1	CADD_HUMAN	P55290 homo sapien
14	42	53.8	886	1	DSC1_MOUSE	P55849 mus musculu
15	41	52.6	794	1	CADC_HUMAN	P55289 homo sapien
16	40	51.3	266	1	PRCB_YEAST	P30657 saccharomyc
17	40	51.3	1627	1	TP2B_CHICK	O42131 gallus gall
18	40	51.3	1612	1	TP2B_CRILO	Q64399 cricetulus
19	40	51.3	1626	1	TP2B_HUMAN	Q02880 homo sapien
20	40	51.3	1612	1	TP2B_MOUSE	Q64511 mus musculu
21	39	50.0	165	1	FTN2_HAEIN	P43708 haemophilus
22	39	50.0	1073	1	ITA6_HUMAN	P23229 homo sapien
23	39	50.0	1073	1	ITA6_MOUSE	Q61739 mus musculu
24	39	50.0	1356	1	KAB7_YEAST	P31374 saccharomyc
25	39	50.0	646	1	VE1_HPV07	Q05133 human papil
26	39	50.0	1748	1	YNR2_YEAST	P53886 saccharomyc
27	38	48.7	887	1	CAD1_CHICK	P08641 gallus gall
28	38	48.7	790	1	CAD6_MOUSE	P97326 mus musculu
29	38	48.7	881	1	DPOL_HPBGS	P03161 ground squi
30	38	48.7	433	1	FTSZ_ARATH	Q42545 arabidopsis
31	38	48.7	273	1	HLPA_HAEIN	P31728 haemophilus
32	38	48.7	456	1	NIFE_CLOPA	P10996 clostridium
33	38	48.7	626	1	PARC_BORBU	O51066 borrelia bu
34	38	48.7	88	1	RS15_MYCFL	P48780 mycoplasma
35	38	48.7	657	1	VE1_HPV18	P06789 human papil
36	38	48.7	94	1	YRHB_ECOLI	P46857 escherichia
37	37	47.4	921	1	AKH_DAUCA	P37142 daucus caro
38	37	47.4	882	1	CAD1_HUMAN	P12830 homo sapien
39	37	47.4	884	1	CAD1_MOUSE	P09803 mus musculu
40	37	47.4	829	1	CAD3_HUMAN	P22223 homo sapien
41	37	47.4	783	1	CAD5_MOUSE	P55284 mus musculu
42	37	47.4	732	1	CADB_CHICK	P33145 gallus gall
43	37	47.4	814	1	CADF_HUMAN	P55291 homo sapien
44	37	47.4	845	1	DPOL_HPBV2	P03159 hepatitis b
45	37	47.4	730	1	DPOL_HPBV4	P12933 hepatitis b

ALIGNMENTS

```
RESULT   1
DSG3_HUMAN
ID   DSG3_HUMAN      STANDARD;      PRT;   999 AA.
AC   P32926;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN   DSG3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92069753.
RA   AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT   "Autoantibodies against a novel epithelial cadherin in pemphigus
RT   vulgaris, a disease of cell adhesion.";
RL   Cell 67:869-877(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC       CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC       DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC       LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC       AGAINST DSG3.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76482; AAA60230.1; -.
DR   PIR; A41088; IJHUG3.
DR   HSSP; P09803; 1EDH.
DR   MIM; 169615; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24     49       POTENTIAL.
FT   CHAIN       50    999       DESMOGLEIN 3.
FT   DOMAIN      50    615       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   616    640       POTENTIAL.
FT   DOMAIN     641    999       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      50    158       CADHERIN 1.
FT   REPEAT     159    268       CADHERIN 2.
FT   REPEAT     269    383       CADHERIN 3.
FT   REPEAT     386 -> 499       CADHERIN 4.
FT   REPEAT     910    935       DESMOGLEIN REPEAT 1.
FT   REPEAT     936    966       DESMOGLEIN REPEAT 2.
FT   CARBOHYD   110    110       POTENTIAL.
```

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match           100.0%;  Score 78;  DB 1;  Length 999;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||||||||||||
Db     97 FGIFVVDKNTGDINI 111


RESULT   2
DSG2_HUMAN
ID   DSG2_HUMAN     STANDARD;      PRT;  1117 AA.
AC   Q14126;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 2 PRECURSOR (HDGC).
GN   DSG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CARCINOMA;
RX   MEDLINE; 94192736.
RA   SCHAEFER S., KOCH P.J., FRANKE W.W.;
RT   "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT   expression catalogue of the desmoglein subfamily of desmosomal
RT   cadherins.";
RL   Exp. Cell Res. 211:391-399(1994).
RN   [2]
RP   SEQUENCE OF 777-1117 FROM N.A.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z26317; CAA81226.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 125671; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24     48       POTENTIAL.
FT   CHAIN       49   1117       DESMOGLEIN 2.
FT   DOMAIN      49    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    633       POTENTIAL.
FT   DOMAIN     634   1117       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      49    159       CADHERIN 1.
FT   REPEAT     160    272       CADHERIN 2.
FT   REPEAT     273    387       CADHERIN 3.
FT   REPEAT     388    502       CADHERIN 4.
FT   REPEAT     880    911       DESMOGLEIN REPEAT 1.
FT   REPEAT     912    941       DESMOGLEIN REPEAT 2.
FT   REPEAT     942    967       DESMOGLEIN REPEAT 3.
FT   REPEAT     968    991       DESMOGLEIN REPEAT 4.
FT   REPEAT     992   1020       DESMOGLEIN REPEAT 5.
FT   REPEAT    1021   1050       DESMOGLEIN REPEAT 6.
FT   CARBOHYD   111    111       POTENTIAL.
FT   CARBOHYD   181    181       POTENTIAL.
FT   CARBOHYD   308    308       POTENTIAL.
FT   CARBOHYD   461    461       POTENTIAL.
FT   CARBOHYD   513    513       POTENTIAL.
SQ   SEQUENCE  1117 AA;  122385 MW;  84D3B898 CRC32;


  Query Match           71.8%;  Score 56;  DB 1;  Length 1117;
  Best Local Similarity  60.0%;  Pred. No. 0.13;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||  :|:||::|:
Db     98 FGIFVFNKDTGELNV 112


RESULT   3
DSG1_BOVIN
ID   DSG1_BOVIN     STANDARD;      PRT;  1043 AA.
AC   Q03763;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
GN   DSG1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANKE W.W.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 44-1043 FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 91168965.
RA   KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA   ZIMBELMANN R., FRANKE W.W.;
RT   "Identification of desmoglein, a constitutive desmosomal
RT   glycoprotein, as a member of the cadherin family of cell adhesion
RT   molecules.";
RL   Eur. J. Cell Biol. 53:1-12(1990).
RN   [3]
RP   REVISIONS, AND SEQUENCE OF 101-123.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA   FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
RN   [4]
RP   SEQUENCE OF 44-493 FROM N.A.
RX   MEDLINE; 91097553.
RA   GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT   "Desmoglein shows extensive homology to the cadherin family of cell
RT   adhesion molecules.";
RL   Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
```

```
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58466; CAA41380.1; -.
DR   EMBL; X57784; CAA40930.1; -.
DR   EMBL; M58165; AAA62709.1; -.
DR   PIR; S14603; IJBOG1.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1043       DESMOGLEIN 1.
FT   DOMAIN       50    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    549    573       POTENTIAL.
FT   DOMAIN      574   1043       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    498       CADHERIN 4.
FT   REPEAT      819    845       DESMOGLEIN REPEAT 1.
FT   REPEAT      846    875       DESMOGLEIN REPEAT 2.
FT   REPEAT      876    905       DESMOGLEIN REPEAT 3.
FT   REPEAT      906    933       DESMOGLEIN REPEAT 4.
FT   REPEAT      934    962       DESMOGLEIN REPEAT 5.
FT   DOMAIN      963   1012       GLY/SER-RICH.
FT   CARBOHYD    110    110
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    496    496       POTENTIAL.
FT   CONFLICT    124    124       I -> V (IN REF. 4).
SQ   SEQUENCE   1043 AA;  112243 MW;  13898584 CRC32;


  Query Match           70.5%;  Score 55;  DB 1;  Length 1043;
  Best Local Similarity  60.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDINI 15
           :||||::: ||:|||
Db      97 YGIFVINQKTGEINI 111


RESULT   4
DSG1_HUMAN
ID   DSG1_HUMAN     STANDARD;      PRT;  1049 AA.
AC   Q02413;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN   DSG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 91271279.
RA   WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOTIS P., POYNTER D.,
RA   ARNEMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA   BUXTON R.S., MAGEE A.I.;
RT   "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT   junctions, is related to the cadherin family of cell adhesion
RT   molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56654; CAA39976.1; -.
DR   PIR; S16906; IJHUG1.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125670; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1049       DESMOGLEIN 1.
FT   DOMAIN       50    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    570       POTENTIAL.
FT   DOMAIN      571   1049       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    497       CADHERIN 4.
FT   REPEAT      813    839       DESMOGLEIN REPEAT 1.
FT   REPEAT      840    869       DESMOGLEIN REPEAT 2.
FT   REPEAT      870    899       DESMOGLEIN REPEAT 3.
FT   REPEAT      900    927       DESMOGLEIN REPEAT 4.
FT   REPEAT      928    956       DESMOGLEIN REPEAT 5.
FT   DOMAIN      969   1019       GLY/SER-RICH.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    180    180       POTENTIAL.
SQ   SEQUENCE   1049 AA;  113715 MW;  FDD79961 CRC32;


  Query Match           70.5%;  Score 55;  DB 1;  Length 1049;
  Best Local Similarity  60.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDINI 15
           :||||::: ||:|||
Db      97 YGIFVINQKTGEINI 111


RESULT   5
DSC1_BOVIN
ID   DSC1_BOVIN     STANDARD;      PRT;   893 AA.
AC   Q01107; Q28095;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
GN   DSC1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
```

```
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
RC   TISSUE=EPIDERMIS;
RX   MEDLINE; 91185414.
RA   COLLINS J.E., LEGAN P.K., KENNY T.P., MACGARVIE J., HOLTON J.L.,
RA   GARROD D.R.;
RT   "Cloning and sequence analysis of desmosomal glycoproteins 2 and 3
RT   (desmocollins): cadherin-like desmosomal adhesion molecules with
RT   heterogeneous cytoplasmic domains.";
RL   J. Cell Biol. 113:381-391(1991).
RN   [2]
RP   SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
RX   MEDLINE; 91239591.
RA   MECHANIC S., RAYNOR K., HILL J.E., COWIN P.;
RT   "Desmocollins form a distinct subset of the cadherin family of cell
RT   adhesion molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4476-4480(1991).
RN   [3]
RP   SEQUENCE OF 133-893 FROM N.A. (1A), AND PARTIAL SEQUENCE.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 92008912.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., SCHMELZ M.,
RA   FRANKE W.W.;
RT   "Amino acid sequence of bovine muzzle epithelial desmocollin derived
RT   from cloned cDNA: a novel subtype of desmosomal cadherins.";
RL   Differentiation 47:29-36(1991).
RN   [4]
RP   SEQUENCE OF 133-155.
RX   MEDLINE; 91115997.
RA   HOLTON J.L., KENNY T.P., LEGAN P.K., COLLINS J.E., KEEN J.N.,
RA   SHARMA R., GARROD D.R.;
RT   "Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal
RT   similarity to calcium-dependent cell-cell adhesion molecules.";
RL   J. Cell Sci. 97:239-246(1990).
RN   [5]
RP   PHOSPHORYLATION.
RX   MEDLINE; 91009551.
RA   PARRISH E.P., MARSTON J.E., MATTEY D.L., MEASURES H.R., VENNING R.,
RA   GARROD D.R.;
RT   "Size heterogeneity, phosphorylation and transmembrane organisation
RT   of desmosomal glycoproteins 2 and 3 (desmocollins) in MDCK cells.";
RL   J. Cell Sci. 96:239-248(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
CC       TO THE KERATINIZATION OF EPITHELIAL TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; 1A OR DG2 (SHOWN HERE) AND 1B OR
CC       DG3; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS AND WEAKLY IN TONGUE PAPILLAE.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- PTM: FORM 1A IS PHOSPHORYLATED ON A SERINE BUT FORM 1B IS NOT.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M67489; AAA30492.1; -.
DR   EMBL; X56966; CAA40286.1; -.
DR   EMBL; X56967; CAA40287.1; -.
DR   EMBL; X56968; CAA40289.1; -.
DR   EMBL; X56968; CAA40288.1; -.
DR   EMBL; X58029; CAA41088.1; -.
DR   PIR; B38456; IJBODF.
DR   PIR; A43838; IJBODE.
DR   PIR; A39377; A39377.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane; Signal;
KW   Repeat; Cytoskeleton; Calcium-binding; Alternative splicing.
FT   SIGNAL       1     29
FT   PROPEP      30    132
FT   CHAIN       133    893       DESMOCOLLIN 1A/1B.
FT   DOMAIN      133    692       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    693    715       POTENTIAL.
FT   DOMAIN      716    893       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      133    240       CADHERIN 1.
FT   REPEAT      241    352       CADHERIN 2.
FT   REPEAT      353    470       CADHERIN 3.
FT   REPEAT      471    574       CADHERIN 4.
FT   REPEAT      575    682       CADHERIN 5.
FT   CARBOHYD    163    163       POTENTIAL.
FT   CARBOHYD    398    398       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
FT   VARIANT     519    519       I -> F.
FT   VARIANT     788    788       Y -> C.
FT   VARSPLIC    829    839       KVYLCGQDEEH -> ESIRGHTLVKN (IN FORM
FT                                1B/DG3).
FT   VARSPLIC    840    893       MISSING (IN FORM 1B/DG3).
FT   CONFLICT    485    485       T -> A (IN REF. 2 AND 3).
SQ   SEQUENCE   893 AA;  99647 MW;  FF5DD78B CRC32;


  Query Match             57.7%;  Score 45;  DB 1;  Length 893;
  Best Local Similarity   53.8%;  Pred. No. 6.8;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDI 13
           | :| ::|:|||| 
Db     180 FNLFFIEKDTGDI 192


RESULT   6
DSC1_HUMAN
ID   DSC1_HUMAN     STANDARD;      PRT;   894 AA.
AC   Q08554;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
GN   DSC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FORESKIN;
RX   MEDLINE; 93283249.
RA   THEIS D.G., KOCH P.J., FRANKE W.W.;
RT   "Differential synthesis of type 1 and type 2 desmocollin mRNAs in
RT   human stratified epithelia.";
RL   Int. J. Dev. Biol. 37:101-110(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FORESKIN;
RA   ZIMBELMANN R.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RX   MEDLINE; 94116981.
RA   KING I.A., ARNEMANN J., SPURR N.K., BUXTON R.S.;
RT   "Cloning of the cDNA (DSC1) coding for human type 1 desmocollin and
RT   its assignment to chromosome 18.";
RL   Genomics 18:185-194(1993).
RN   [4]
```

```
RP   SEQUENCE OF 135-151 AND 283-292.
RX   MEDLINE; 91323543.
RA   KING I.A., MAGEE A.I., REES D.A., BUXTON R.S.;
RT   "Keratinization is associated with the expression of a new protein
RT   related to the desmosomal cadherins DGII/III.";
RL   FEBS Lett. 286:9-12(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
CC       TO THE KERATINIZATION OF EPITHELIAL TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; 1A OR DG2 (SHOWN HERE) AND 1B OR
CC       DG3; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN EPIDERMIS, LESS IN LYMPH
CC       NODE AND TONGUE.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z34522; CAA84279.1; -.
DR   EMBL; Z34522; CAA84278.1; -.
DR   EMBL; X72925; CAA51428.1; -.
DR   EMBL; X72925; CAA51429.1; -.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125643; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding; Alternative splicing.
FT   SIGNAL       1    29       POTENTIAL.
FT   PROPEP      30   134       POTENTIAL.
FT   CHAIN      135   894       DESMOCOLLIN 1A/1B.
FT   DOMAIN     135   691       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   692   714       POTENTIAL.
FT   DOMAIN     715   894       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     135   242       CADHERIN 1.
FT   REPEAT     243   354       CADHERIN 2.
FT   REPEAT     355   471       CADHERIN 3.
FT   REPEAT     472   575       CADHERIN 4.
FT   REPEAT     576   682       CADHERIN 5.
FT   CARBOHYD   165   165       POTENTIAL.
FT   CARBOHYD   546   546       POTENTIAL.
FT   VARSPLIC   830   840       KVYLCGQDEEH -> ESIRGHTLIKN (IN FORM
FT                              1B/DG3).
FT   VARSPLIC   841   894       MISSING (IN FORM 1B/DG3).
FT   CONFLICT   132   132       T -> S (IN REF. 3).
SQ   SEQUENCE   894 AA;  100044 MW;  A287BCA9 CRC32;


  Query Match             57.7%;  Score 45;  DB 1;  Length 894;
  Best Local Similarity   53.8%;  Pred. No. 6.8;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDI 13
           | :| ::|:||||
Db     182 FNLFYIEKDTGDI 194


RESULT   7
CAD5_PIG
ID   CAD5_PIG       STANDARD;      PRT;   782 AA.
AC   O02840;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).
GN   CDH5.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KILSHAW P.J.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC       IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC       COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
CC       ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
CC       BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13919; CAA74225.1; -.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1    22       POTENTIAL.
FT   PROPEP      23    44       POTENTIAL.
FT   CHAIN       45   782       VASCULAR ENDOTHELIAL-CADHERIN.
FT   DOMAIN      45   592       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   593   619       POTENTIAL.
FT   DOMAIN     620   782       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      45   148       CADHERIN 1.
FT   REPEAT     149   255       CADHERIN 2.
FT   REPEAT     256   370       CADHERIN 3.
FT   REPEAT     371   475       CADHERIN 4.
FT   REPEAT     476   592       CADHERIN 5.
FT   DOMAIN     736   751       SER-RICH.
FT   CARBOHYD    58    58       POTENTIAL.
FT   CARBOHYD   154   154       POTENTIAL.
FT   CARBOHYD   360   360       POTENTIAL.
FT   CARBOHYD   440   440       POTENTIAL.
FT   CARBOHYD   522   522       POTENTIAL.
FT   CARBOHYD   534   534       POTENTIAL.
SQ   SEQUENCE   782 AA;  87546 MW;  007F70E0 CRC32;


  Query Match             55.1%;  Score 43;  DB 1;  Length 782;
  Best Local Similarity   63.6%;  Pred. No. 13;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :| ||:||||:
Db      89 VFRVDENTGDV 99


RESULT   8
CAD6_HUMAN
ID   CAD6_HUMAN     STANDARD;      PRT;   790 AA.
AC   P55285;
DT   01-OCT-1996 (Rel. 34, Created)
```

```
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CADHERIN-6 PRECURSOR (KIDNEY-CADHERIN) (K-CADHERIN).
GN   CDH6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95262134.
RA   SHIMOYAMA Y., GOTOH M., TERASAKI T., KITAJIMA M., HIROHASHI S.;
RT   "Isolation and sequence analysis of human cadherin-6 complementary
RT   DNA for the full coding sequence and its expression in human
RT   carcinoma cells.";
RL   Cancer Res. 55:2206-2211(1995).
RN   [2]
RP   SEQUENCE OF 377-790 FROM N.A.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE; 91283540.
RA   SUZUKI S., SANO K., TANIHARA H.;
RT   "Diversity of the cadherin family: evidence for eight new cadherins
RT   in nervous tissue.";
RL   Cell Regul. 2:261-270(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN, CEREBELLUM, AND
CC       KIDNEY. LUNG, PANCREAS, AND GASTRIC MUCOSA SHOW A WEAK EXPRESSION.
CC       ALSO EXPRESSED IN CERTAIN LIVER AND KIDNEY CARCINOMAS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D31784; BAA06562.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 603007; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19     53       POTENTIAL.
FT   CHAIN        54    790       CADHERIN-6.
FT   DOMAIN       54    615       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    616    636       POTENTIAL.
FT   DOMAIN      637    790       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       54    159       CADHERIN 1.
FT   REPEAT      160    268       CADHERIN 2.
FT   REPEAT      269    383       CADHERIN 3.
FT   REPEAT      384    486       CADHERIN 4.
FT   REPEAT      487    608       CADHERIN 5.
FT   CARBOHYD     49     49       POTENTIAL.
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    399    399       POTENTIAL.
FT   CARBOHYD    437    437       POTENTIAL.
FT   CARBOHYD    455    455       POTENTIAL.
FT   CARBOHYD    536    536       POTENTIAL.
FT   CONFLICT    421    421       V -> I (IN REF. 2).
FT   CONFLICT    425    425       T -> I (IN REF. 2).
SQ   SEQUENCE   790 AA;  88308 MW;  FC415E1E CRC32;
```

```
  Query Match           55.1%;  Score 43;  DB 1;  Length 790;
  Best Local Similarity  54.5%;  Pred. No. 13;
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT   9
CAD6_RAT
ID   CAD6_RAT       STANDARD;      PRT;   789 AA.
AC   P55280;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CADHERIN-6 PRECURSOR (KIDNEY-CADHERIN) (K-CADHERIN).
GN   CDH6 OR KCAD.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ACI; TISSUE=KIDNEY;
RX   MEDLINE; 94243827.
RA   XIANG Y.Y., TANAKA M., SUZUKI M., IGARASHI H., KIYOKAWA E., NAITO Y.,
RA   OHTAWARA Y., SHEN Q., SUGIMURA H., KINO I.;
RT   "Isolation of complementary DNA encoding K-cadherin, a novel rat
RT   cadherin preferentially expressed in fetal kidney and kidney
RT   carcinoma.";
RL   Cancer Res. 54:3034-3041(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY AND BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D25290; BAA04975.1; -.
DR   HSSP; P15116; 1NCI.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19     53       POTENTIAL.
FT   CHAIN        54    789       CADHERIN-6.
FT   DOMAIN       54    615       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    616    636       POTENTIAL.
FT   DOMAIN      637    789       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       54    159       CADHERIN 1.
FT   REPEAT      160    268       CADHERIN 2.
FT   REPEAT      269    383       CADHERIN 3.
FT   REPEAT      384    486       CADHERIN 4.
FT   REPEAT      487    608       CADHERIN 5.
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    399    399       POTENTIAL.
FT   CARBOHYD    437    437       POTENTIAL.
FT   CARBOHYD    455    455       POTENTIAL.
FT   CARBOHYD    536    536       POTENTIAL.
SQ   SEQUENCE   789 AA;  88340 MW;  403CCCB3 CRC32;
```

```
  Query Match           55.1%;  Score 43;  DB 1;  Length 789;
  Best Local Similarity  54.5%;  Pred. No. 13;
```

```
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT  10
CADA_CHICK
ID   CADA_CHICK     STANDARD;      PRT;   789 AA.
AC   P79995;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CADHERIN-10 PRECURSOR.
GN   CDH10.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND RETINA;
RX   MEDLINE; 97358546.
RA   FUSHIMI D., ARNDT K., TAKEICHI M., REDIES C.;
RT   "Cloning and expression analysis of cadherin-10 in the CNS of the
RT   chicken embryo.";
RL   Dev. Dyn. 209:269-285(1997).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000512; BAA19130.1; -.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   PROPEP      23     54       POTENTIAL.
FT   CHAIN       55    789       CADHERIN-10.
FT   DOMAIN      55    606       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   607    634       POTENTIAL.
FT   DOMAIN     635    789       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      55    160       CADHERIN 1.
FT   REPEAT     161    269       CADHERIN 2.
FT   REPEAT     270    384       CADHERIN 3.
FT   REPEAT     385    487       CADHERIN 4.
FT   REPEAT     488    606       CADHERIN 5.
FT   CARBOHYD   256    256       POTENTIAL.
FT   CARBOHYD   438    438       POTENTIAL.
FT   CARBOHYD   456    456       POTENTIAL.
FT   CARBOHYD   534    534       POTENTIAL.
SQ   SEQUENCE   789 AA;  88392 MW;  D44464F0 CRC32;


  Query Match           55.1%;  Score 43;  DB 1;  Length 789;
  Best Local Similarity   50.0%;  Pred. No. 13;
  Matches    6;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           :|::|: ||||:
Db     101 LFIIDEKTGDIH 112


RESULT  11
CADE_HUMAN
ID   CADE_HUMAN     STANDARD;      PRT;   790 AA.
AC   Q13634;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CADHERIN-14 PRECURSOR.
GN   CDH14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97184182.
RA   SHIBATA T., SHIMOYAMA Y., GOTOH M., HIROHASHI S.;
RT   "Identification of human cadherin-14, a novel neurally specific type
RT   II cadherin, by protein interaction cloning.";
RL   J. Biol. Chem. 272:5236-5240(1997).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U59325; AAB02933.1; -.
DR   HSSP; P15116; 1NCI.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   PROPEP      25     53       POTENTIAL.
FT   CHAIN       54    790       CADHERIN-14.
FT   DOMAIN      54    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   609    636       POTENTIAL.
FT   DOMAIN     637    790       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      54    159       CADHERIN 1.
FT   REPEAT     160    268       CADHERIN 2.
FT   REPEAT     269    383       CADHERIN 3.
FT   REPEAT     384    486       CADHERIN 4.
FT   REPEAT     487    608       CADHERIN 5.
FT   CARBOHYD    36     36       POTENTIAL.
FT   CARBOHYD   255    255       POTENTIAL.
FT   CARBOHYD   455    455       POTENTIAL.
FT   CARBOHYD   536    536       POTENTIAL.
SQ   SEQUENCE   790 AA;  88072 MW;  4596258F CRC32;


  Query Match           55.1%;  Score 43;  DB 1;  Length 790;
  Best Local Similarity   58.3%;  Pred. No. 13;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           ||::|  ||||:
Db     100 IFIIDDTTGDIH 111


RESULT  12
```

```
YGA3_YEAST
ID   YGA3_YEAST     STANDARD;      PRT;   566 AA.
AC   P53197;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL 62.8 KD TRP-ASP REPEATS CONTAINING PROTEIN IN PMC1-TFG2
DE   INTERGENIC REGION.
GN   YGL003C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   HEBLING U., HOFMANN B., DELIUS H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z72525; CAA96703.1; -.
DR   PFAM; PF00400; WD40; 4.
DR   PROSITE; PS00678; WD_REPEATS; 1.
KW   Hypothetical protein; Repeat; WD repeat.
FT   DOMAIN       32     38       POLY-SER.
FT   REPEAT      300    330       WD1.
FT   REPEAT      342    370       WD2.
FT   REPEAT      383    413       WD3.
FT   REPEAT      425    458       WD4.
FT   REPEAT      469    501       WD5.
FT   REPEAT      513    543       WD6.
SQ   SEQUENCE   566 AA;  62821 MW;  65B1272B CRC32;


  Query Match           55.1%;  Score 43;  DB 1;  Length 566;
  Best Local Similarity  63.6%;  Pred. No. 9;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           ||: | ||||:
Db     285 IFLTDNNTGDV 295


RESULT  13
CADD_HUMAN
ID   CADD_HUMAN     STANDARD;      PRT;   713 AA.
AC   P55290;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   T-CADHERIN PRECURSOR (TRUNCATED-CADHERIN) (CADHERIN-13) (H-CADHERIN)
DE   (HEART-CADHERIN).
GN   CDH13 OR CDHH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 95073006.
RA   TANIHARA H., SANO K., HEIMARK R.L., ST JOHN T., SUZUKI S.;
RT   "Cloning of five human cadherins clarifies characteristic features of
RT   cadherin extracellular domain and provides further evidence for two
RT   structurally different types of cadherin.";
RL   Cell Adhes. Commun. 2:15-26(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96266329.
RA   LEE S.W.;
RT   "H-cadherin, a novel cadherin with growth inhibitory functions and
RT   diminished expression in human breast cancer.";
RL   Nat. Med. 2:776-782(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98407802.
RA   SATO M., MORI Y., SAKURADA A., FUJIMURA S., HORII A.;
RT   "The H-cadherin (CDH13) gene is inactivated in human lung cancer.";
RL   Hum. Genet. 103:96-101(1998).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34058; AAA35624.1; -.
DR   EMBL; U59289; AAB18912.1; -.
DR   EMBL; U59288; AAB18911.1; -.
DR   EMBL; AB001103; BAA32411.1; -.
DR   EMBL; AB001090; BAA32411.1; JOINED.
DR   EMBL; AB001091; BAA32411.1; JOINED.
DR   EMBL; AB001092; BAA32411.1; JOINED.
DR   EMBL; AB001093; BAA32411.1; JOINED.
DR   EMBL; AB001094; BAA32411.1; JOINED.
DR   EMBL; AB001095; BAA32411.1; JOINED.
DR   EMBL; AB001096; BAA32411.1; JOINED.
DR   EMBL; AB001097; BAA32411.1; JOINED.
DR   EMBL; AB001098; BAA32411.1; JOINED.
DR   EMBL; AB001099; BAA32411.1; JOINED.
DR   EMBL; AB001100; BAA32411.1; JOINED.
DR   EMBL; AB001101; BAA32411.1; JOINED.
DR   EMBL; AB001102; BAA32411.1; JOINED.
DR   HSSP; P15116; 1NCI.
DR   MIM; 601364; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Calcium-binding; Repeat;
KW   GPI-anchor; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23    138       POTENTIAL.
FT   CHAIN       139    693       T-CADHERIN.
FT   PROPEP      694    713       REMOVED IN MATURE FORM (POTENTIAL).
FT   REPEAT      139    245       CADHERIN 1.
FT   REPEAT      246    363       CADHERIN 2.
FT   REPEAT      364    477       CADHERIN 3.
FT   REPEAT      478    583       CADHERIN 4.
FT   REPEAT      583    693       CADHERIN 5.
FT   CARBOHYD     52     52       POTENTIAL.
FT   CARBOHYD     86     86       POTENTIAL.
FT   CARBOHYD    382    382       POTENTIAL.
FT   CARBOHYD    500    500       POTENTIAL.
FT   CARBOHYD    530    530       POTENTIAL.
FT   CARBOHYD    598    598       POTENTIAL.
FT   CARBOHYD    638    638       POTENTIAL.
FT   CARBOHYD    671    671       POTENTIAL.
FT   LIPID       693    693       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   713 AA;  78286 MW;  8BC381E5 CRC32;

  Query Match           53.8%;  Score 42;  DB 1;  Length 713;
  Best Local Similarity  42.9%;  Pred. No. 17;
```

```
  Matches    6;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 GIFVVDKNTGDINI 15
           ||| :::||| .:::
Db     186 GIFRINENTGSVSV 199


RESULT  14
DSC1_MOUSE
ID   DSC1_MOUSE      STANDARD;      PRT;   886 AA.
AC   P55849;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DESMOCOLLIN 1A/1B PRECURSOR.
GN   DSC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=SKIN;
RX   MEDLINE; 96420658.
RA   KING I.A., O'BRIEN T.J., BUXTON R.S.;
RT   "Expression of the 'skin-type' desmosomal cadherin DSC1 is closely
RT   linked to the keratinization of epithelial tissues during mouse
RT   development.";
RL   J. Invest. Dermatol. 107:531-538(1996).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
CC       TO THE KERATINIZATION OF EPITHELIAL TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: FORMS 1A AND 1B ARE PRODUCED BY ALTERNATIVE
CC       SPLICING OF THE SAME GENE.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X97986; CAA66628.1; -.
DR   EMBL; X97986; CAA66629.1; -.
DR   HSSP; P09803; 1EDH.
DR   MGD; MGI:109173; DSC1.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat; Alternative splicing.
FT   SIGNAL       1    29       POTENTIAL.
FT   PROPEP      30   134       POTENTIAL.
FT   CHAIN      135   886       DESMOCOLLIN 1A/1B.
FT   DOMAIN     135   691       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   692   714       POTENTIAL.
FT   DOMAIN     715   886       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     135   242       CADHERIN 1.
FT   REPEAT     243   354       CADHERIN 2.
FT   REPEAT     355   471       CADHERIN 3.
FT   REPEAT     472   575       CADHERIN 4.
FT   REPEAT     576   682       CADHERIN 5.
FT   CARBOHYD   130   130       POTENTIAL.
FT   CARBOHYD   165   165       POTENTIAL.
FT   CARBOHYD   546   546       POTENTIAL.
FT   CARBOHYD   613   613       POTENTIAL.
FT   VARSPLIC   822   832       KVYLCGQAEEH -> ESIRGHTLIKN (IN FORM 1B).
FT   VARSPLIC   840   886       MISSING (IN FORM 1B).
SQ   SEQUENCE   886 AA;  98953 MW;  6C3C94BA CRC32;


  Query Match             53.8%;  Score 42;  DB 1;  Length 886;
  Best Local Similarity   46.2%;  Pred. No. 21;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDI 13
           : :| ::|:||||
Db     182 YNLFYIEKDTGDI 194


RESULT  15
CADC_HUMAN
ID   CADC_HUMAN      STANDARD;      PRT;   794 AA.
AC   P55289;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   BRAIN-CADHERIN PRECURSOR (BR-CADHERIN) (CADHERIN-12) (N-CADHERIN 2)
DE   (CADHERIN, NEURAL TYPE, 2).
GN   CDH12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 95073006.
RA   TANIHARA H., SANO K., HEIMARK R.L., ST JOHN T., SUZUKI S.;
RT   "Cloning of five human cadherins clarifies characteristic features of
RT   cadherin extracellular domain and provides further evidence for two
RT   structurally different types of cadherin.";
RL   Cell Adhes. Commun. 2:15-26(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN CORTEX;
RX   MEDLINE; 95249541.
RA   SELIG S., BRUNO S., SCHARF J.M., WANG C.H., VITALE E., GILLIAM T.C.,
RA   KUNKEL L.M.;
RT   "Expressed cadherin pseudogenes are localized to the critical region
RT   of the spinal muscular atrophy gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3702-3706(1995).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L34057; AAA35623.1; -.
DR   EMBL; L33477; AAB48539.1; -.
DR   HSSP; P09803; 1EDH.
DR   MIM; 600562; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1    23       POTENTIAL.
FT   PROPEP      24    54       POTENTIAL.
FT   CHAIN       55   794       BRAIN-CADHERIN.
FT   DOMAIN      55   609       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM    610    637       POTENTIAL.
FT   DOMAIN      638    794       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       55    160       CADHERIN 1.
FT   REPEAT      161    269       CADHERIN 2.
FT   REPEAT      270    384       CADHERIN 3.
FT   REPEAT      385    487       CADHERIN 4.
FT   REPEAT      488    609       CADHERIN 5.
FT   CARBOHYD    256    256       POTENTIAL.
FT   CARBOHYD    456    456       POTENTIAL.
FT   CARBOHYD    537    537       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
FT   CONFLICT     68     68       V -> M (IN REF. 2).
FT   CONFLICT    349    349       E -> D (IN REF. 2).
FT   CONFLICT    416    416       G -> S (IN REF. 2).
FT   CONFLICT    644    644       H -> D (IN REF. 2).
FT   CONFLICT    733    733       I -> Y (IN REF. 2).
FT   CONFLICT    761    761       A -> T (IN REF. 2).
SQ   SEQUENCE   794 AA;  88274 MW;  78BF3F7E CRC32;


  Query Match            52.6%;  Score 41;  DB 1;  Length 794;
  Best Local Similarity  50.0%;  Pred. No. 28;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           :| :|: ||||:
Db     101 VFTIDETTGDIH 112


Search completed: January  7, 2000, 13:25:37
Job time: 389 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:18 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-2
Perfect score: 78
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	67	85.9	993	11	O35902	O35902 mus musculu
2	51	65.4	785	13	Q90763	Q90763 gallus gall
3	43	55.1	214	2	Q55701	Q55701 synechocyst
4	43	55.1	1329	5	O76356	O76356 caenorhabdi
5	43	55.1	950	13	O93508	O93508 brachydanio
6	43	55.1	790	13	Q90762	Q90762 gallus gall
7	42	53.8	1069	4	O60245	O60245 homo sapien
8	42	53.8	1072	4	O60246	O60246 homo sapien
9	42	53.8	1200	4	O60247	O60247 homo sapien
10	42	53.8	3380	5	Q24292	Q24292 drosophila
11	41.5	53.2	148	12	O37171	O37171 heliothis a
12	41	52.6	4307	5	Q19319	Q19319 caenorhabdi
13	40	51.3	598	1	O29907	O29907 archaeoglob
14	40	51.3	1033	2	O33373	O33373 neisseria g
15	40	51.3	813	11	Q63315	Q63315 rattus norv
16	40	51.3	694	11	Q63561	Q63561 rattus norv
17	40	51.3	1035	13	O57537	O57537 xenopus lae
18	39	50.0	814	4	O75281	O75281 homo sapien
19	39	50.0	803	4	O75283	O75283 homo sapien
20	39	50.0	816	4	O75284	O75284 homo sapien
21	39	50.0	391	5	Q19859	Q19859 caenorhabdi
22	39	50.0	205	5	Q27742	Q27742 plasmodium
23	39	50.0	801	11	Q9Z0M3	Q9z0m3 mus musculu
24	39	50.0	73	13	Q06275	Q06275 xenopus lae
25	39	50.0	790	13	Q91838	Q91838 xenopus lae
26	39	50.0	1087	13	Q91778	Q91778 xenopus lae
27	38.5	49.4	798	2	Q50288	Q50288 mycoplasma
28	38	48.7	245	1	Q58723	Q58723 methanococc
29	38	48.7	1055	2	P95632	P95632 rickettsia
30	38	48.7	1058	2	P95661	P95661 rickettsia
31	38	48.7	828	2	P72602	P72602 synechocyst
32	38	48.7	497	2	P74864	P74864 salmonella
33	38	48.7	341	2	O86281	O86281 lactococcus
34	38	48.7	914	4	O95151	O95151 homo sapien
35	38	48.7	831	5	P91388	P91388 caenorhabdi
36	38	48.7	423	10	O65875	O65875 pisum sativ
37	38	48.7	797	11	Q63418	Q63418 rattus norv
38	38	48.7	274	12	Q84181	Q84181 human papil
39	38	48.7	1616	12	Q9YKD6	Q9ykd6 tomato mosa
40	37.5	48.1	444	2	Q9Z9U5	Q9z9u5 bacillus sp
41	37	47.4	359	1	O27477	O27477 methanobact
42	37	47.4	1060	2	P95585	P95585 rickettsia
43	37	47.4	1060	2	P95659	P95659 rickettsia
44	37	47.4	977	2	O25308	O25308 helicobacte
45	37	47.4	1616	12	O41352	O41352 tobacco mos

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 85.9%; Score 67; DB 11; Length 993;
Best Local Similarity 86.7%; Pred. No. 0.0046;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 FGIFVVDKNTGDINI 15
          ||||||| | |||||
Db     97 FGIFVVDPNNGDINI 111
```

RESULT 2
Q90763
ID Q90763 PRELIMINARY; PRT; 785 AA.
AC Q90763;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CHICKEN CADHERIN-7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.

```
RC   STRAIN=WHITE LEGHORN; TISSUE=BRAIN;
RX   MEDLINE; 95309115.
RA   NAKAGAWA S., TAKEICHI M.;
RT   "Neural crest cell-cell adhesion controlled by sequential and
RT   subpopulation-specific expression of novel cadherins.";
RL   Development 121:1321-1332(1995).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D42150; BAA07721.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   785 AA;  87171 MW;  F27A6881 CRC32;


  Query Match             65.4%;  Score 51;  DB 13;  Length 785;
  Best Local Similarity   66.7%;  Pred. No. 1.7;
  Matches    8;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 IFVVDKNTGDIN 14
          ||::|:|||||:
Db     94 IFIIDENTGDIH 105


RESULT   3
Q55701
ID   Q55701      PRELIMINARY;      PRT;   214 AA.
AC   Q55701;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   HYPOTHETICAL 24.5 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 96127529.
RA   KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA   SUGIURA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D64000; BAA10231.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   214 AA;  24539 MW;  CDEA936D CRC32;


  Query Match             55.1%;  Score 43;  DB 2;  Length 214;
  Best Local Similarity   42.9%;  Pred. No. 8.8;
  Matches    6;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDIN 14
          ||: ::| :|||:|:
Db     49 FGVLMIDPSTGEIS 62


RESULT   4
O76356
ID   O76356      PRELIMINARY;      PRT;  1329 AA.
AC   O76356;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   C45G7.6 PROTEIN.
GN   C45G7.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DANTE M., WAMSLEY P.;
RT   "The sequence of C. elegans cosmid C45G7.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF067611; AAC19184.1; -.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1329 AA;  146518 MW;  B161D39E CRC32;


  Query Match             55.1%;  Score 43;  DB 5;  Length 1329;
  Best Local Similarity   72.7%;  Pred. No. 65;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 IFVVDKNTGDI 13
          || |||::|||
Db   1288 IFAVDKDSGDI 1298


RESULT   5
O93508
ID   O93508      PRELIMINARY;      PRT;   950 AA.
AC   O93508;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PARAXIAL PROTOCADHERIN.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC   Cyprinoidea; Cyprinidae; Rasborinae; Danio.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   YAMAMOTO A., DE ROBERTIS E.M.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF042191; AAC62385.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   950 AA;  105381 MW;  96B102F0 CRC32;

  Query Match             55.1%;  Score 43;  DB 13;  Length 950;
  Best Local Similarity   63.6%;  Pred. No. 45;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :| :|||||:|
Db     615 LFSIDKNTGEI 625


RESULT   6
Q90762
ID   Q90762      PRELIMINARY;      PRT;   790 AA.
AC   Q90762;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CHICKEN CADHERIN-6B.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=BRAIN;
RX   MEDLINE; 95309115.
RA   NAKAGAWA S., TAKEICHI M.;
RT   "Neural crest cell-cell adhesion controlled by sequential and
RT   subpopulation-specific expression of novel cadherins.";
RL   Development 121:1321-1332(1995).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D42149; BAA07720.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   790 AA;  88659 MW;  DE01FC76 CRC32;

  Query Match             55.1%;  Score 43;  DB 13;  Length 790;
  Best Local Similarity   54.5%;  Pred. No. 37;
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT   7
O60245
ID   O60245      PRELIMINARY;      PRT;  1069 AA.
AC   O60245;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PCDH7 (BH-PCDH)A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006755; BAA25194.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1069 AA;  116104 MW;  F1732B30 CRC32;

  Query Match             53.8%;  Score 42;  DB 4;  Length 1069;
  Best Local Similarity   58.3%;  Pred. No. 75;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 GIFVVDKNTGDI 13
           ||| :| ::|||
Db     580 GIFAIDPDSGDI 591


RESULT   8
O60246
ID   O60246      PRELIMINARY;      PRT;  1072 AA.
AC   O60246;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PCDH7 (BH-PCDH)B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006756; BAA25195.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 6.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1072 AA;  116463 MW;  A3DF367C CRC32;

  Query Match             53.8%;  Score 42;  DB 4;  Length 1072;
  Best Local Similarity   58.3%;  Pred. No. 75;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 GIFVVDKNTGDI 13
           ||| :| ::|||
Db     580 GIFAIDPDSGDI 591


RESULT   9
O60247
ID   O60247      PRELIMINARY;      PRT;  1200 AA.
AC   O60247;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PCDH7 (BH-PCDH)C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006757; BAA25196.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1200 AA;  130337 MW;  56F1CD33 CRC32;

  Query Match             53.8%;  Score 42;  DB 4;  Length 1200;
  Best Local Similarity   58.3%;  Pred. No. 85;
```

```
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 GIFVVDKNTGDI 13
           ||| :| ::|||
Db     533 GIFAIDPDSGDI 544


RESULT  10
Q24292
ID   Q24292      PRELIMINARY;      PRT;  3380 AA.
AC   Q24292;'
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ADHERIN.
GN   DACHSOUS.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95324813.
RA   CLARK H.F., BRENTRUP D., SCHNEITZ K., BIEBER A., GOODMAN C., NOLL M.;
RT   "Dachsous encodes a member of the cadherin superfamily that controls
RT   imaginal disc morphogenesis in Drosophila.";
RL   Genes Dev. 9:1530-1542(1995).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; L08811; AAA79329.1; -.
DR   FLYBASE; FBgn0000497; ds.
DR   PFAM; PF00028; cadherin; 25.
DR   PROSITE; PS00232; CADHERIN; 18.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   3380 AA;  366356 MW;  7D83FAC3 CRC32;


  Query Match           53.8%;  Score 42;  DB 5;  Length 3380;
  Best Local Similarity  58.3%;  Pred. No. 2.7e+02;
  Matches    7;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GIFVVDKNTGDI 13
           ||| :|::||:|
Db     713 GIFRIDRSTGEI 724


RESULT  11
O37171
ID   O37171      PRELIMINARY;      PRT;   148 AA.
AC   O37171;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE   17K ORF.
OS   Heliothis armigera entomopoxvirus (HaEPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SRISKANTHA A., OSBORNE R.J., DALL D.J.;
RL   J. Gen. Virol. 0:0-0(1997).
DR   EMBL; AF017791; AAB96622.1; -.
SQ   SEQUENCE   148 AA;  17488 MW;  A945DB09 CRC32;


  Query Match           53.2%;  Score 41.5;  DB 12;  Length 148;
  Best Local Similarity  42.1%;  Pred. No. 10;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    5;  Gaps    1;

Qy       1 FGIF-----VVDKNTGDIN 14
           :|:|     ::||||| |
Db      17 YGLFGDFKLIIDKNTGYFN 35


RESULT  12
Q19319
ID   Q19319      PRELIMINARY;      PRT;  4307 AA.
AC   Q19319; Q19785; Q21606;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   F25F2.2 PROTEIN.
GN   F25F2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   AINSCOUGH R.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; Z35662; CAA84721.1; -.
DR   EMBL; Z35599; CAA84721.1; JOINED.
DR   EMBL; Z34802; CAA84721.1; JOINED.
DR   EMBL; Z35599; CAA84661.1; -.
DR   EMBL; Z35662; CAA84661.1; JOINED.
DR   EMBL; Z34802; CAA84661.1; JOINED.
DR   EMBL; Z34802; CAA84339.1; -.
DR   EMBL; Z35599; CAA84339.1; JOINED.
DR   EMBL; Z35662; CAA84339.1; JOINED.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 9.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   4307 AA;  478958 MW;  3486FC46 CRC32;


  Query Match           52.6%;  Score 41;  DB 5;  Length 4307;
  Best Local Similarity  72.7%;  Pred. No. 5.1e+02;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :| || |||||
Db    3467 LFRVDPNTGDI 3477


RESULT  13
O29907
ID   O29907      PRELIMINARY;      PRT;   598 AA.
AC   O29907;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1999 (TrEMBLrel. 11, Last annotation update)
DE   PROBABLE TUNGSTEN-CONTAINING ALDEHYDE FERREDOXIN OXIDOREDUCTASE 3
DE   (EC 1.2.7.-) (AOR-3).
GN   AF0340.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic,
RT   sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- PATHWAY: PYROGLYCOLYTIC PATHWAY.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
```

```
CC   -!- COFACTOR: EACH SUBUNIT CONTAINS A 4FE4S CLUSTER AND A TUNGSTEN
CC       COFACTOR
CC       (CONSISTING OF 2 MOLYBDOPTERIN MOLECULES COORDINATED TO A
CC       TUNGSTEN BY A TOTAL OF 4 SULFUR LIGANDS) (BY SIMILARITY).
CC   -!- DOMAIN: CAN BE DIVIDED INTO 3 DOMAINS.
CC       DOMAIN I WHICH FORMS A BASE ON WHICH THE SADDLE-LIKE TUNGSTEN
CC       PTERIN COFACTOR SITS AND, CONTAINS THE MAGNESIUM BINDING SITES.
CC       DOMAINS II AND III SURROUND THE OPPOSITE SURFACE OF THE TUNGSTEN
CC       COFACTOR AND ARE INVOLVED IN INTERACTIONS WITH THE DIFFERENT
CC       METAL CENTERS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AOR/FOR/GAPOR/CAR/HVOR FAMILY.
DR   EMBL; AE001081; AAB90895.1; -.
DR   TIGR; AF0340; -.
DR   PFAM; PF01314; AFOR; 1.
KW   Hypothetical protein; Oxidoreductase; Iron-sulfur; 4Fe-4S;
KW   Multigene family; Tungsten.
FT   DOMAIN        1    216       DOMAIN I.
FT   DOMAIN      217    413       DOMAIN II.
FT   DOMAIN      414    598       DOMAIN III.
FT   METAL       291    291       IRON-SULFUR (BY SIMILARITY).
FT   METAL       294    294       IRON-SULFUR (BY SIMILARITY).
FT   METAL       298    298       IRON-SULFUR (BY SIMILARITY).
FT   METAL       493    493       IRON-SULFUR (BY SIMILARITY).
SQ   SEQUENCE   598 AA;  65666 MW;  909BAB0E CRC32;


  Query Match             51.3%;  Score 40;  DB 1;  Length 598;
  Best Local Similarity   40.0%;  Pred. No. 86;
  Matches    6;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDINI 15
           | ::: :|  || ||
Db     355 FAMYLTEKGVGDFNI 369


RESULT  14
O33373
ID   O33373      PRELIMINARY;      PRT;  1033 AA.
AC   O33373;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PILC PROTEIN.
GN   PILC.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GC-653;
RX   MEDLINE; 98129088.
RA   BAECKMAN M., KAELLSTROEM H., JONSSON A.B.;
RT   "The phase-variable pilus-associated protein PilC is commonly
RT   expressed in clinical isolates of Neisseria gonorrhoeae, and shows
RT   sequence variability among strains.";
RL   Microbiology 144:149-156(1998).
DR   EMBL; AJ001121; CAA04547.1; -.
SQ   SEQUENCE   1033 AA;  112936 MW;  B0593053 CRC32;


  Query Match             51.3%;  Score 40;  DB 2;  Length 1033;
  Best Local Similarity   46.7%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDINI 15
           :|||  || || : :
Db     804 YGIFDDDKGTGTVKV 818


RESULT  15
Q63315
ID   Q63315      PRELIMINARY;      PRT;  813 AA.
AC   Q63315;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   LONG TYPE PB-CADHERIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=BRAIN;
RX   MEDLINE; 96212232.
RA   SUGIMOTO K., HONDA S., YAMAMOTO T., UEKI T., MONDEN M., KAJI A.,
RA   MATSUMOTO K., NAKAMURA T.;
RT   "Molecular cloning and characterization of a newly identified member
RT   of the cadherin family, PB-cadherin.";
RL   J. Biol. Chem. 271:11548-11556(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D83348; BAA11894.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   813 AA;  87978 MW;  6E4D9F1D CRC32;


  Query Match             51.3%;  Score 40;  DB 11;  Length 813;
  Best Local Similarity   58.3%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           ||::|: ||||:
Db     106 IFLIDELTGDIH 117


Search completed: January 12, 2000, 23:15:19
Job time: 182 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:36 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-4
Perfect score: 76
Sequence: 1 TPMFLLSRNTGEVRT 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	76	100.0	999	1	R30742	Human pemphigus vu
2	76	100.0	614	1	W07908	Pemphigus vulgaris
3	76	100.0	15	1	W04844	Self epitope of de
4	76	100.0	15	1	W64816	Desmoglein-3 206-2
5	76	100.0	15	1	W78815	Desmoglein 3 prote
6	64	84.2	778	1	W15489	Pemphigus foliaceu
7	48	63.2	616	1	R86867	Rat protocadherin
8	43	56.6	560	1	W13009	Segment of desmoso
9	43	56.6	263	1	W13010	Segment of desmoso
10	39	51.3	249	1	W38513	S. pneumoniae poss
11	39	51.3	835	1	W80699	S. pneumoniae cati
12	38	50.0	586	1	R07998	Asparagine synthet
13	38	50.0	583	1	R07999	Asparagine synthet
14	37	48.7	407	1	P94879	ORF incorporated w
15	37	48.7	218	1	W62750	Streptococcus pneu
16	36.5	48.0	387	1	R49142	Fowlpox virus prot
17	36.5	48.0	387	1	R58857	Fowlpox virus (FPV
18	36.5	48.0	387	1	W10726	Homology vector 44
19	36	47.4	878	1	R55060	Sequence of human
20	36	47.4	878	1	R85487	Human E-cadherin p
21	36	47.4	484	1	R99735	Retinoid X recepto
22	36	47.4	451	1	R99739	Retinoid X recepto
23	36	47.4	469	1	W03448	Farnesoid-activate
24	36	47.4	191	1	W04552	Retroviral reverse
25	35	46.1	609	1	P30206	Sequence encoded b
26	35	46.1	219	1	R38092	nodB. Isolated nod
27	35	46.1	1148	1	R43671	M.leprae rpoB gene
28	35	46.1	284	1	W27703	E. coli AIDA-I pro
29	35	46.1	472	1	W40072	Human retinoid rec
30	35	46.1	20	1	W40251	Human wild-type E-
31	35	46.1	555	1	W84154	Human desaturase e
32	35	46.1	608	1	W84155	Human desaturase e
33	35	46.1	746	1	W84156	Human desaturase e
34	35	46.1	746	1	W85135	A desaturase enzym
35	35	46.1	555	1	W85133	A desaturase enzym
36	35	46.1	608	1	W85134	A desaturase enzym
37	34.5	45.4	163	1	W38654	S. pneumoniae aspa
38	34	44.7	610	1	P93578	Extracellular neut
39	34	44.7	609	1	R21414	Vibrolysin, a prot
40	34	44.7	792	1	R59765	Extracellular neut
41	34	44.7	609	1	W86187	Vibriolysin, usefu
42	33.5	44.1	720	1	W19266	Lactobacillus amyl
43	33	43.4	478	1	R04881	Recombinant elasta
44	33	43.4	337	1	R12995	GAP-DH. Promoter o
45	33	43.4	30	1	R93960	Peptide combining

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 76; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTGEVRT 15
||||||||||||||||
Db 206 TPMFLLSRNTGEVRT 220

RESULT 2
W07908
ID W07908 standard; protein; 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphis vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucuous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose

CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 76; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 6.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTGEVRT 15
||||||||||||||||
Db 205 TPMFLLSRNTGEVRT 219

RESULT 3
W04844
ID W04844 standard; peptide; 15 AA.
AC W04844;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN WO9627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1; Page 40; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-myslin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 206-220)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTGEVRT 15
||||||||||||||||
Db 1 TPMFLLSRNTGEVRT 15

RESULT 4
W64816
ID W64816 standard; peptide; 15 AA.
AC W64816;
DT 29-SEP-1998 (first entry)
DE Desmoglein-3 206-220.
KW Desmoglein; DG; gene therapy; pemphigus vulgaris; microparticle;
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PD 21-JUL-1998.
PF 22-JAN-1997; 787547.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS;
DR WPI; 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PT expression of proteins e.g. in gene therapy
PS Disclosure; Column 4; 42pp; English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TPMFLLSRNTGEVRT 15
||||||||||||||||
Db 1 TPMFLLSRNTGEVRT 15

RESULT 5
W78815
ID W78815 standard; peptide; 15 AA.
AC W78815;
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 206-220.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
OS Synthetic.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998; U01499.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR WPI; 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure; Page 8; 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less that 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (i) a
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment

```
CC   of a naturally-occurring protein from an infectious agent which infects
CC   a mammal; (2) a peptide having a length and sequence which permits it to
CC   bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC   peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC   to W78897 are peptide fragments for use in the present invention. The
CC   MPs are highly effective vehicles for the delivery of polynucleotides
CC   into phagocytic cells. They can be used for gene therapy, e.g. for
CC   treating genetic diseases, infections or tumours or for downregulating
CC   an immune response.
SQ   Sequence   15 AA;


  Query Match             100.0%;  Score 76;  DB 1;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEVRT 15
          |||||||||||||||
Db      1 TPMFLLSRNTGEVRT 15


RESULT   6
W15489
ID   W15489 standard; Protein; 778 AA.
AC   W15489;
DT   17-JUN-1997  (first entry)
DE   Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW   Pemphigus foliaceus; autoantibody; constant region; IgG;
KW   extracellular region; antigen; hinge portion; skin;
KW   dermatitis herpetiformis; fusion protein; detection; ss.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1. .545
FT                   /note= "Pemphigus foliaceus antigen protein"
PN   J09077800-A.
PD   25-MAR-1997.
PF   12-SEP-1995; 260899.
PR   12-SEP-1995; JP-260899.
PA   (NISH/) NISHIKAWA T.
DR   WPI; 97-241758/22.
DR   P-PSDB; T66428.
PT   Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT   through the hinge region used to treat pemphigus foliaceus
PS   Claim 1; Page 10-12; 17pp; Japanese.
CC   This sequence represents a fused protein recognised by pemphigus
CC   foliaceus patient autoantibody which comprises the constant region
CC   of IgG linked to the extracellular region of pemphigus foliaceus
CC   antigen protein through the hinge portion.  Pemphigus foliaceus is
CC   a chronic, generalised, vesicular and scaling skin eruption similar
CC   to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC   protein is useful to treat pemphigus foliaceus.  The antigen is
CC   especially administered through an adsorbent upon which the fusion
CC   protein is immobilised via a carrier. The fusion protein is also
CC   useful for detecting pemphigus foliaceus antibodies which is useful
CC   in immunodiagnosis.  The fusion protein has little or no side effects.
SQ   Sequence   778 AA;


  Query Match             84.2%;  Score 64;  DB 1;  Length 778;
  Best Local Similarity   66.7%;  Pred. No. 0.0012;
  Matches   10;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEVRT 15
          :|||:::|||||:||
Db    206 SPMFIINRNTGEIRT 220


RESULT   7
R86867
ID   R86867 standard; Protein; 616 AA.
AC   R86867;
DT   27-AUG-1996  (first entry)
DE   Rat protocadherin pc5.
KW   Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW   catenin; therapy.
OS   Rattus rattus.
PN   WO9600289-A1.
PD   04-JAN-1996.
PF   26-JUN-1995; U08071.
PR   27-JUN-1994; US-268161.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 96-068873/07.
DR   N-PSDB; T03574.
PT   Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT   pc5 - involved in cell-cell adhesion and regulation activities
PS   Claim 17; Page 130-132; 146pp; English.
CC   R86865-R86867 represent the sequences for three protocadherins.  This
CC   sequence represents the rat protocadherin pc5.  These sequences are
CC   related to cadherin, and possess cell adhesive ability.  Cadherins are
CC   glycosylated integral membrane proteins that are involved in cell-cell
CC   adhesion.  Cadherins are composed of an N-terminal extracellular domain
CC   which consists of 5 unique subdomains, a membrane spanning domain, and a
CC   C-terminal cytoplasmic domain.  The cytoplasmic domain interacts with the
CC   cytoskeleton through catenins and other cytoskeleton associated proteins.
CC   The cytoplasmic domain is not present in all cadherins, but in those
CC   which possess it, it is essential for the cadherins adhesive function.
CC   The cadherins which do not possess a cytoplasmic domain appear to
CC   function via a different method from those with a cytoplasmic domain.
CC   These sequences were isolated using primers 1 and 2 (see T03575 and
CC   T03576)  The proteins may have regulatory functions in the cell, as well
CC   as the cell-cell adhesive properties.  Antibodies produced against these
CC   sequences are useful for modulating the binding activity of these
CC   protocadherins, and can be used therapeutically.
SQ   Sequence   616 AA;


  Query Match             63.2%;  Score 48;  DB 1;  Length 616;
  Best Local Similarity   61.5%;  Pred. No. 0.7;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEV 13
          ||:| |: ||||:
Db     85 TPLFQLNENTGEI 97


RESULT   8
W13009
ID   W13009 standard; protein; 560 AA.
AC   W13009;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 7; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
```

```
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
CC   carcinomas.
SQ   Sequence   560 AA;


  Query Match             56.6%;  Score 43;  DB 1;  Length 560;
  Best Local Similarity   50.0%;  Pred. No. 5.1;
  Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 PMFLLSRNTGEVRT 15
          |:| |:::|||: |
Db    160 PVFYLNKDTGEIYT 173


RESULT   9
W13010
ID   W13010 standard; protein; 263 AA.
AC   W13010;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 9; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
CC   carcinomas.
SQ   Sequence   263 AA;


  Query Match             56.6%;  Score 43;  DB 1;  Length 263;
  Best Local Similarity   50.0%;  Pred. No. 2.2;
  Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 PMFLLSRNTGEVRT 15
          |:| |:::|||: |
Db     58 PVFYLNKDTGEIYT 71


RESULT  10
W38513
ID   W38513 standard; Protein; 249 AA.
AC   W38513;
DT   06-NOV-1998  (first entry)
DE   S. pneumoniae possible cation transporting ATPase PACL.
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis.
OS   Streptococcus pneumoniae.
PN   WO9743303-A1.
PD   20-NOV-1997.
PF   14-MAY-1997; U07950.
PR   14-MAY-1996; US-017670.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Stodola RK;
DR   WPI; 98-008793/01.
DR   N-PSDB; T98578.
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT   diagnosing anti-microbial agents for treatment of bacterial
PT   infections
PS   Claim 12; Page 296; 483pp; English.
CC   This represents a Streptococcus pneumoniae protein that is a possible
CC   cation transporting ATPase, and is encoded by a DNA  of the invention.
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC   invention can be used to identify compounds which interact with and
CC   inhibit or activate the activity of the proteins. Antagonists can be
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic
CC   immunisation. They can also be used to induce an immunological response
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC   of the encoding nucleic acids in a vector adequate to produce antibody
CC   and/or T cell immune responses to protect the animal from disease. The
CC   proteins can also be used to identify antimicrobial compounds which are
CC   capable of inhibiting their bioactivity. In particular the proteins of
CC   the invention can be used to prevent adhesion of bacteria to mammalian
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to
CC   block protein-mediated mammalian cell invasion, and to block the normal
CC   progression of pathogenesis in infections initiated other than by the
CC   implantation of in-dwelling devices or other surgical techniques.
SQ   Sequence   249 AA;


  Query Match             51.3%;  Score 39;  DB 1;  Length 249;
  Best Local Similarity   60.0%;  Pred. No. 11;
  Matches    9;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSRNTGEVRT 15
          |  :||| || || |
Db    201 TIQYLLSANTAEVLT 215


RESULT  11
W80699
ID   W80699 standard; Protein; 835 AA.
AC   W80699;
DT   24-DEC-1998  (first entry)
DE   S. pneumoniae cation transporting ATPase.
KW   Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW   virulence; antibody; infection; detection; treatment; hypothetical;
KW   cell wall biosynthetic, external target; minimal gene set protein.
OS   Streptococcus pneumoniae.
PN   WO9826072-A1.
PD   18-JUN-1998.
PF   09-DEC-1997; U22578.
PR   13-DEC-1996; US-036281.
PA   (ELIL ) LILLY & CO ELI.
PI   Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI   Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
PI   Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI   Young Bellido ML;
DR   WPI; 98-348529/30.
DR   N-PSDB; V65261.
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT   for evaluating gene expression, and identification of virulence
PT   genes
PS   Claim 3; Pages 288-291; 333pp; English.
CC   This sequence represents a Streptococcus pneumoniae cation transporting
CC   ATPase. The invention provides DNA sequences (V65201 to V65304) from
CC   the Streptococcus pneumoniae genome and corresponding protein sequences
```

```
CC   (W80605 to W80728). The protein sequences are classified as hypothetical,
CC   cell wall biosynthetic, external target, or minimal gene set proteins. A
CC   recombinant host containing a vector comprising any of the above nucleic
CC   acids can be used for the recombinant expression of the proteins. The
CC   invention also provides a DNA chip having arrayed on it at least 15 base
CC   pair fragment of any one or more of these DNA sequences. The DNA chip can
CC   be used methods for evaluating gene expression in S. pneumoniae and for
CC   identifying virulence genes in S. pneumoniae. Antibodies that selectively
CC   bind to the above proteins or peptide fragments can be used to treat
CC   S. pneumoniae infection. The antibodies can also be used to detect
CC   S. pneumoniae cells.
SQ   Sequence   835 AA;


  Query Match             51.3%;  Score 39;  DB 1;  Length 835;
  Best Local Similarity   60.0%;  Pred. No. 41;
  Matches    9;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSRNTGEVRT 15
           |  :||| || || |
Db     622 TIQYLLSANTAEVLT 636


RESULT  12
R07998
ID   R07998 standard; protein; 586 AA.
AC   R07998;
DT   26-FEB-1991  (first entry)
DE   Asparagine synthetase AS1.
KW   asparagine synthetase; transgenic plant; herbicide resistance;
KW   drought tolerance; nitrogen fixation; pea.
OS   Pisum sativum.
FH   Key             Location/Qualifiers
FT   region          116. .128
FT                   /label= 80% homology with human AS1 sequence
FT   region          218. .243
FT                   /label= 80% homology with human AS1
FT   region          340. .348
FT                   /label= 80% homology with human AS1
FT   region          352. .360
FT                   /label= 80% homology with human AS1
FT   region          392. .401
FT                   /label= 80% homology with human AS1
FT   region          486. .500
FT                   /label= 80% homology with human AS1
FT   binding_site    1. .4
FT                   /label= putative glutamine binding site
PN   WO9013633-A.
PD   15-NOV-1990.
PF   02-MAY-1990; U02443.
PR   03-MAY-1989; US-347302.
PR   26-APR-1990; US-514816.
PA   (UYRO-) ROCKFELLER UNIV.
PI   Coruzzi GM, Tsai FY;
DR   WPI; 90-361471/48.
DR   N-PSDB; Q06598.
PT   Plant asparagine synthetase - includes DNA expression system and
PT   transgenic plants
PS   Disclosure; Fig 2A; 91pp; English.
CC   The DNA sequence encoding this protein was isolated from cDNA
CC   clones selected from a pea nodule cDNA library from the "Sparkle"
CC   variety of P.sativum. Human AS cDNA was used as a probe.
CC   The protein is produced by expression vectors containing the AS1
CC   coding sequence. Recombinant AS can be used to engineer herbicide
CC   resistance, as a dominant selectable marker, to select novel
CC   herbicides or compounds useful for synchronising plant cells in
CC   culture, etc.
CC   See also Q06599, Q06622 and Q06623.
SQ   Sequence   586 AA;


  Query Match             50.0%;  Score 38;  DB 1;  Length 586;
  Best Local Similarity   87.5%;  Pred. No. 43;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     323 TPMFLMSR 330


RESULT  13
R07999
ID   R07999 standard; protein; 583 AA.
AC   R07999;
DT   26-FEB-1991  (first entry)
DE   Asparagine synthetase AS2.
KW   asparagine synthetase; transgenic plant; herbicide resistance;
KW   drought tolerance; nitrogen fixation; pea.
OS   Pisum sativum.
FH   Key             Location/Qualifiers
FT   binding_site    1. .4
FT                   /label= glutamine binding site
FT                   /note= "conserved in human AS and pea AS1 and 2."
PN   WO9013633-A.
PD   15-NOV-1990.
PF   02-MAY-1990; U02443.
PR   03-MAY-1989; US-347302.
PR   26-APR-1990; US-514816.
PA   (UYRO-) ROCKFELLER UNIV.
PI   Coruzzi GM, Tsai FY;
DR   WPI; 90-361471/48.
DR   N-PSDB; Q06599.
PT   Plant asparagine synthetase - includes DNA expression system and
PT   transgenic plants
PS   Disclosure; Fig 2B; 91pp; English.
CC   The DNA sequence encoding this protein was isolated from cDNA
CC   clones selected from a pea nodule cDNA library from the "Sparkle"
CC   variety of P.sativum. Pea AS1 cDNA was used as a probe.
CC   The protein is produced by expression vectors containing the AS2
CC   coding sequence. Recombinant AS can be used to engineer herbicide
CC   resistance, as a dominant selectable marker, to select novel
CC   herbicides or compounds useful for synchronising plant cells in
CC   culture, etc.
CC   A comparison of pea AS and human AS polypeptides reveals an overall
CC   homology of 47% at the amino acid level, c.f. 86% between AS1 and
CC   AS2. There are several regions of high local homology (greater than
CC   80%) shared between the pea AS and human AS polypeptides.
CC   See also Q06598, Q06622 and Q06623.
SQ   Sequence   583 AA;


  Query Match             50.0%;  Score 38;  DB 1;  Length 583;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     322 TPMFLMSR 329


RESULT  14
P94879
ID   P94879 standard; protein; 407 AA.
AC   P94879;
DT   10-MAR-1993  (revised)
DT   02-JUL-1990  (first entry)
DE   ORF incorporated within the HIS4 gene.
KW   Yeast promoter; yeast pheromones; FUS-1; BIK-1; HIS-4.
OS   Saccharomyces cerevisiae.
PN   WO8810308-A.
PD   29-DEC-1988.
PF   23-JUN-1988; 02129.
PR   24-JUN-1987; US-066078.
PA   (WHIT-) Whitehead Inst.
PI   Fink GR, Trueheart J, Elion EA;
DR   WPI; 89-023850/03.
```

```
DR   N-PSDB; N93100.
PT   New DNA fragment contg. protein encoding gene and yeast promoter -
PT   controlled by mating pheromone allowing efficient and
PT   regulatable expression.
PS   Claim 5; Fig 4; 51pp; English.
CC   ORF lies between the FUS 1 or BIK 1 promoter, BIK 1 running 3' to 5'.
CC   At least one polypeptide may be promoted within a high-copy vector
CC   induced by a-factor for alpha cells, alpha-factor for a-cells using this
CC   promoter system.
SQ   Sequence   407 AA;


  Query Match             48.7%;  Score 37;  DB 1;  Length 407;
  Best Local Similarity   57.1%;  Pred. No. 44;
  Matches    8;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEVRT 15
           ||    ||| | ||
Db     366 PMSTEDRNTAETRT 379


RESULT  15
W62750
ID   W62750 standard; Protein; 218 AA.
AC   W62750;
DT   09-NOV-1998  (first entry)
DE   Streptococcus pneumoniae polypeptide.
KW   Polypeptide; ORF; open reading frame; infection; bacterial;
KW   streptococcal; bacteremia; diagnosis; prophylaxis.
OS   Streptococcus pneumoniae.
PN   WO9823631-A1.
PD   04-JUN-1998.
PF   24-NOV-1997; U21976.
PR   27-NOV-1996; US-031879.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI   Reid RH, Zarfos PN;
DR   WPI; 98-322654/28.
PT   Streptococcus pneumoniae polynucleotides - useful for developing
PT   products for diagnosis, prevention and treatment of infections e.g.
PT   pneumonia, bacteremia, meningitis or endocarditis
PS   Claim 5; Page 32; 181pp; English.
CC   The sequence is that of a Streptococcal polypeptide.
CC   The polypeptide can potentially be used for the diagnosis and
CC   prevention of bacterial infections, especially SP infection.
CC   It may be used for the treatment of diseases such as otitis media,
CC   conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC   empyema, endocarditis or infection of the cerebrospinal fluid.
SQ   Sequence   218 AA;


  Query Match             48.7%;  Score 37;  DB 1;  Length 218;
  Best Local Similarity   38.9%;  Pred. No. 22;
  Matches    7;  Conservative    6;  Mismatches    1;  Indels    4;  Gaps    1;

Qy       2 PMF----LLSRNTGEVRT 15
           |:|    |::: |||::|
Db      44 PIFVTFRLINKETGEIKT 61


Search completed: January  6, 2000, 15:40:37
Job time: 1684 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:21 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-4
Perfect score: 76
Sequence: 1 TPMFLLSRNTGEVRT 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	62	81.6	993	11	O35902	O35902 mus musculu
2	48	63.2	820	4	O15039	O15039 homo sapien
3	47	61.8	336	5	Q09661	Q09661 caenorhabdi
4	43.5	57.2	291	2	Q04824	Q04824 salmonella
5	40	52.6	452	1	O29889	O29889 archaeoglob
6	39	51.3	393	2	O08395	O08395 mycobacteri
7	39	51.3	163	2	O84262	O84262 chlamydia t
8	39	51.3	215	5	O44017	O44017 entamoeba h
9	38	50.0	810	4	O75279	O75279 homo sapien
10	38	50.0	1115	4	O95846	O95846 homo sapien
11	38	50.0	585	10	O65329	O65329 elaeagnus u
12	38	50.0	586	10	Q40328	Q40328 medicago sa
13	38	50.0	586	10	P93618	P93618 vicia faba
14	38	50.0	581	10	P93167	P93167 glycine max
15	38	50.0	586	10	O24483	O24483 medicago sa
16	38	50.0	948	10	O81211	O81211 arabidopsis
17	38	50.0	578	10	Q9ZST7	Q9zst7 arabidopsis
18	38	50.0	578	10	Q9ZST6	Q9zst6 arabidopsis
19	38	50.0	584	10	Q9ZS23	Q9zs23 phaseolus v
20	37	48.7	231	2	P71027	P71027 bacillus su
21	37	48.7	211	2	O83378	O83378 treponema p
22	37	48.7	426	2	Q9ZMS6	Q9zms6 helicobacte
23	37	48.7	557	3	O42902	O42902 schizosacch
24	37	48.7	398	3	O60145	O60145 schizosacch
25	37	48.7	407	3	O94079	O94079 saccharomyc
26	37	48.7	816	4	O75284	O75284 homo sapien
27	37	48.7	234	10	O04149	O04149 arabidopsis
28	37	48.7	367	10	Q9ZSJ0	Q9zsj0 arabidopsis
29	37	48.7	593	12	O10288	O10288 orgyia pseu
30	36	47.4	466	2	P77286	P77286 escherichia
31	36	47.4	932	4	O60330	O60330 homo sapien
32	36	47.4	807	4	O75277	O75277 homo sapien
33	36	47.4	814	4	O75281	O75281 homo sapien
34	36	47.4	803	4	O75283	O75283 homo sapien
35	36	47.4	878	4	Q15855	Q15855 homo sapien
36	36	47.4	414	5	Q21405	Q21405 caenorhabdi
37	36	47.4	643	5	Q18290	Q18290 caenorhabdi
38	36	47.4	1417	10	O48697	O48697 arabidopsis
39	36	47.4	483	10	O49927	O49927 pisum sativ
40	36	47.4	928	11	O08963	O08963 rattus norv
41	36	47.4	308	11	O35100	O35100 rattus norv
42	36	47.4	469	11	Q62735	Q62735 rattus norv
43	36	47.4	379	11	Q60643	Q60643 mus musculu
44	36	47.4	484	11	Q60641	Q60641 mus musculu
45	36	47.4	914	11	O88192	O88192 mus musculu

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 81.6%; Score 62; DB 11; Length 993;
Best Local Similarity 92.3%; Pred. No. 0.0024;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 MFLLSRNTGEVRT 15
           |||:|||||||||
Db     208 MFLISRNTGEVRT 220
```

RESULT 2
O15039
ID O15039 PRELIMINARY; PRT; 820 AA.
AC O15039;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE KIAA0327 PROTEIN.
GN KIAA0327.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97349984.
RA   NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA   TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB002325; BAA20785.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   820 AA;  89864 MW;  3B6ECA92 CRC32;


  Query Match             63.2%;  Score 48;  DB 4;  Length 820;
  Best Local Similarity   61.5%;  Pred. No. 0.84;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSRNTGEV 13
           ||:| |: ||||:
Db     289 TPLFQLNENTGEI 301


RESULT   3
Q09661
ID   Q09661      PRELIMINARY;      PRT;   336 AA.
AC   Q09661;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 37.7 KD PROTEIN ZK177.6 IN CHROMOSOME II.
GN   ZK177.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   ANDERSON K.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: WEAK, TO YEAST PROTEIN CDC20.
DR   EMBL; U21321; AAB36970.1; -.
DR   WORMPEP; ZK177.6; CE02095.
DR   PFAM; PF00400; WD40; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   336 AA;  37668 MW;  2DC77B26 CRC32;


  Query Match             61.8%;  Score 47;  DB 5;  Length 336;
  Best Local Similarity   57.1%;  Pred. No. 0.49;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEVRT 15
           |||  :  :|||:|||
Db     187 PMFTIDEHTGQVRT 200


RESULT   4
Q04824
ID   Q04824      PRELIMINARY;      PRT;   291 AA.
AC   Q04824;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   ORF6 PROTEIN PRECURSOR.
OS   Salmonella typhimurium.
OG   Plasmid 90 kb virulence.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE; 93316852.
RA   FRIEDRICH M.J., KINSEY N.E., VILA J., KADNER R.J.;
RT   "Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence
RT   plasmid of Salmonella typhimurium: the presence of fimbrial
RT   biosynthetic genes.";
RL   Mol. Microbiol. 8:543-558(1993).
DR   EMBL; L08613; AAC36963.1; -.
KW   Plasmid; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    291       ORF6 PROTEIN.
SQ   SEQUENCE   291 AA;  30683 MW;  C4D5ADDF CRC32;


  Query Match             57.2%;  Score 43.5;  DB 2;  Length 291;
  Best Local Similarity   80.0%;  Pred. No. 1.9;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 TPMFLLSRNTGEVRT 15
           | || ||| ||||||
Db     128 TLMFSLSR-TGEVRT 141


RESULT   5
O29889
ID   O29889      PRELIMINARY;      PRT;   452 AA.
AC   O29889;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   DNA REPAIR PROTEIN RAD25.
GN   AF0358.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA   KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA   RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA   FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA   KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA   PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA   OVERBEEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA   COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA   SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA   MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the hyperthermophilic,
RT   sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001080; AAB90879.1; -.
DR   TIGR; AF0358; -.
DR   PFAM; PF00270; DEAD; 1.
DR   PFAM; PF00271; helicase_C; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   452 AA;  51768 MW;  615C80A2 CRC32;


  Query Match             52.6%;  Score 40;  DB 1;  Length 452;
  Best Local Similarity   72.7%;  Pred. No. 14;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 LLSRNTGEVRT 15
           |:|| |||| |
Db     430 LISRGTGEVNT 440


RESULT   6
O08395
```

```
ID   O08395      PRELIMINARY;      PRT;   393 AA.
AC   O08395;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CITRATE SYNTHASE (EC 4.1.3.7).
GN   GLTA3.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   DEVLIN K., CHURCHER C.M.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
CC   -!- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF
CC       OXIDATIVE METABOLISM.
CC   -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC       OXALOACETATE.
CC   -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC   -!- SIMILARITY: TO OTHER CITRATE SYNTHASES.
DR   EMBL; Z95585; CAB09042.1; -.
DR   PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR   PFAM; PF00285; citrate_synt; 1.
KW   Lyase; Tricarboxylic acid cycle.
SQ   SEQUENCE   393 AA;  42969 MW;  8B50A8D3 CRC32;


  Query Match             51.3%;  Score 39;  DB 2;  Length 393;
  Best Local Similarity   63.6%;  Pred. No. 19;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSRNTG 11
           ||:|::|| ||
Db     350 TPIFVMSRITG 360


RESULT   7
O84262
ID   O84262      PRELIMINARY;      PRT;   163 AA.
AC   O84262;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 18.9 KD PROTEIN.
GN   CT260.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001299; AAC67853.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   163 AA;  18875 MW;  80D724D1 CRC32;


  Query Match             51.3%;  Score 39;  DB 2;  Length 163;
  Best Local Similarity   58.3%;  Pred. No. 7.3;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEV 13
           |:| : || |||
Db      43 PLFFVIRNEGEV 54


RESULT   8
O44017
ID   O44017      PRELIMINARY;      PRT;   215 AA.
AC   O44017;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ERD2 GENE.
GN   ERD2.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK-9;
RA   SANCHEZ-LOPEZ R., GAMA-CASTRO S., RAMOS M.A., MERINO E., LIZARDI P.M.,
RA   ALAGON A.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ002138; CAA05206.1; -.
DR   PFAM; PF00810; ER_lumen_recept; 1.
SQ   SEQUENCE   215 AA;  25724 MW;  28FF1C82 CRC32;


  Query Match             51.3%;  Score 39;  DB 5;  Length 215;
  Best Local Similarity   66.7%;  Pred. No. 9.8;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PMFLLSRNTGEV 13
           |  || |:||||
Db     135 PQILLLRSTGEV 146


RESULT   9
O75279
ID   O75279      PRELIMINARY;      PRT;   810 AA.
AC   O75279;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KIAA0345-LIKE 3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KIMMERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
RA   KADNER K., MIGUEL T., MILLER C., PITLUCK S., POLLARD M., ROJESKI H.,
RA   SUBRAMANIAN S., MARTIN C.H.;
RT   "Sequencing of human chromosome 5.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   RICKE D.O.;
```

```
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AC005609; AAC34316.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   810 AA;  88391 MW;  9F4E3C7D CRC32;

  Query Match             50.0%;  Score 38;  DB 4;  Length 810;
  Best Local Similarity   58.3%;  Pred. No. 64;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 MFLLSRNTGEVR 14
           :| | :| ||||
Db     291 LFTLDQNNGEVR 302


RESULT  10
O95846
ID   O95846      PRELIMINARY;      PRT;  1115 AA.
AC   O95846;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   EUKARYOTIC TRANSLATION INITIATION FACTOR 2 ALPHA KINASE PEK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 99150360.
RA   SHI Y., AN J., LIANG J., HAYES S.E., SANDUSKY G.E., STRAMM L.E.,
RA   YANG N.N.;
RT   "Characterization of a Mutant Pancreatic eIF-2alpha Kinase, PEK, and
RT   Co-localization with Somatostatin in Islet Delta Cells.";
RL   J. Biol. Chem. 274:5723-5730(1999).
DR   EMBL; AF110146; AAD19961.1; -.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Kinase; Initiation factor.
SQ   SEQUENCE   1115 AA;  125146 MW;  5BB6FCC8 CRC32;

  Query Match             50.0%;  Score 38;  DB 4;  Length 1115;
  Best Local Similarity   53.3%;  Pred. No. 90;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSRNTGEVRT 15
           |||   :|:||:| |
Db     972 TPMPAYARHTGQVGT 986


RESULT  11
O65329
ID   O65329      PRELIMINARY;      PRT;  585 AA.
AC   O65329;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ASPARAGINE SYNTHETASE.
GN   AS.
OS   Elaeagnus umbellata.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Rosales; Elaeagnaceae; Elaeagnus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ROOT NODULE;
RA   KIM H.-B., AN C.-S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061740; AAC16325.1; -.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PFAM; PF00310; GATase_2; 1.
SQ   SEQUENCE   585 AA;  65878 MW;  8A12777D CRC32;

  Query Match             50.0%;  Score 38;  DB 10;  Length 585;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     322 TPMFLMSR 329


RESULT  12
Q40328
ID   Q40328      PRELIMINARY;      PRT;  586 AA.
AC   Q40328;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ASPARAGINE SYNTHETASE.
OS   Medicago sativa (Alfalfa).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Medicago.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SARANAC;
RA   GANTT S.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L40327; AAB48058.1; -.
DR   MENDEL; 9036; MEDsa;1042;1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PFAM; PF00310; GATase_2; 1.
SQ   SEQUENCE   586 AA;  66462 MW;  962C3A3F CRC32;

  Query Match             50.0%;  Score 38;  DB 10;  Length 586;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TPMFLLSR 8
           |||||:||
Db     323 TPMFLMSR 330


RESULT  13
P93618
ID   P93618      PRELIMINARY;      PRT;  586 AA.
AC   P93618;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ASPARAGINE SYNTHETASE (EC 6.3.5.4) (ASPARAGINE SYNTHASE
DE   (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
DE   (GLUTAMINE-HYDROLYSING)).
GN   VFAS1.
OS   Vicia faba (Broad bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Vicia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KLEINE THUERINGER; TISSUE=ROOT NODULE;
RA   KUESTER H., ALBUS U., FRUEHLING M., TCHETKOVA S.A., TIKHONOVITCH I.A.,
RA   PUEHLER A., PERLICK A.M.;
RL   Plant Sci. 124:89-95(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
```

```
DR   EMBL; Z72354; CAA96526.1; -.
DR   MENDEL; 11022; VICfa;1042;1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PFAM; PF00310; GATase_2; 1.
KW   Ligase.
SQ   SEQUENCE   586 AA;  66236 MW;  AD8BD927 CRC32;


  Query Match             50.0%;  Score 38;  DB 10;  Length 586;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          |||||:||
Db    323 TPMFLMSR 330


RESULT  14
P93167
ID   P93167      PRELIMINARY;      PRT;   581 AA.
AC   P93167;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ASPARAGINE SYNTHETASE 2 (EC 6.3.5.4) (ASPARAGINE SYNTHASE
DE   (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
DE   (GLUTAMINE-HYDROLYSING)).
GN   AS2.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Glycine.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CENTURY; TISSUE=MATURE LEAF;
RX   MEDLINE; 97188563.
RA   HUGHES C.A., BEARD H., MATTHEWS B.F.;
RT   "Molecular cloning and expression of two cDNAs encoding asparagine
RT   synthetase in soybean.";
RL   Plant Mol. Biol. 33:301-311(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC       DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
DR   EMBL; U77678; AAC49613.1; -.
DR   MENDEL; 8141; GLYma;1042;1.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PFAM; PF00310; GATase_2; 1.
KW   Ligase.
SQ   SEQUENCE   581 AA;  65609 MW;  90AF9E2F CRC32;


  Query Match             50.0%;  Score 38;  DB 10;  Length 581;
  Best Local Similarity   87.5%;  Pred. No. 44;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          |||||:||
Db    322 TPMFLMSR 329


RESULT  15
O24483
ID   O24483      PRELIMINARY;      PRT;   586 AA.
AC   O24483;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ASPARAGINE SYNTHETASE.
OS   Medicago sativa (Alfalfa).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Medicago.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97432147.
RA   SHI L., TWARY S.N., YOSHIOKA H., GREGERSON R.G., MILLER S.S.,
RA   SAMAC D.A., GANTT J.S., UNKEFER P.J., VANCE C.P.;
RT   "Nitrogen assimilation in alfalfa: isolation and characterization of
RT   an asparagine synthetase gene showing enhanced expression in root
RT   nodules and dark-adapted leaves.";
RL   Plant Cell 9:1339-1356(1997).
DR   EMBL; U89923; AAB81011.1; -.
DR   PFAM; PF00733; Asn_synthase; 1.
DR   PFAM; PF00310; GATase_2; 1.
SQ   SEQUENCE   586 AA;  66448 MW;  63C8B6BA CRC32;


  Query Match             50.0%;  Score 38;  DB 10;  Length 586;
  Best Local Similarity   87.5%;  Pred. No. 45;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TPMFLLSR 8
          |||||:||
Db    323 TPMFLMSR 330


Search completed: January 12, 2000, 23:15:22
Job time: 185 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:42 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-5
Perfect score: 88
Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	100.0	999	1	IJHUG3	desmoglein 3 precu
2	74	84.1	1049	1	IJHUG1	desmoglein 1 precu
3	74	84.1	1043	1	IJBOG1	desmoglein 1 precu
4	52	59.1	832	2	S55396	LI-cadherin - huma
5	51	58.0	2163	2	T15276	hypothetical prote
6	49	55.7	863	1	IJBODC	desmocollin 2a pre
7	49	55.7	809	1	IJBODD	desmocollin 2b pre
8	48	54.5	5147	1	IJFFTM	cadherin-related t
9	48	54.5	784	1	IJHUC5	cadherin 5 precurs
10	47	53.4	896	2	I45858	desmocollin - bovi
11	47	53.4	3343	2	S44887	ZK112.7 protein -
12	46	52.3	1117	2	S38673	desmoglein 2 - hum
13	46	52.3	793	2	D38992	cadherin 8 - human
14	46	52.3	790	2	G02678	cadherin-14 - huma
15	46	52.3	889	2	T09055	protocadherin 68 -
16	45	51.1	901	1	IJHUDA	desmocollin 3a pre
17	45	51.1	847	1	IJHUDB	desmocollin 3b pre
18	45	51.1	794	2	I59372	cadherin 12 - huma
19	45	51.1	805	2	E70474	translation initia
20	45	51.1	656	2	T10664	serine/threonine-s
21	45	51.1	827	2	A53954	LI-cadherin precur
22	44	50.0	884	1	IJMSCE	E-cadherin precurs
23	44	50.0	884	2	S34438	uvomorulin - mouse
24	44	50.0	790	2	I37016	cadherin-6 - human
25	44	50.0	789	2	I52701	K-cadherin - rat
26	44	50.0	785	2	I50180	cadherin-7 - chick
27	44	50.0	790	2	I50178	cadherin-6B - chic
28	44	50.0	790	2	I51638	F-cadherin - Afric
29	44	50.0	807	2	T12177	potassium channel
30	44	50.0	787	2	S68699	potassium channel
31	44	50.0	4351	2	T00252	MEGF1 protein - ra
32	44	50.0	682	2	T06106	proliferating-cell
33	44	50.0	620	2	T15273	hypothetical prote
34	44	50.0	3051	2	S42373	hypothetical prote
35	44	50.0	3097	2	T00021	DN-cadherin - frui
36	44	50.0	829	2	I46536	Ksp-cadherin - rab
37	44	50.0	235	2	E70378	DNA replication pr
38	43	48.9	796	2	A38992	cadherin 11 precur
39	43	48.9	796	2	I48277	cadherin-11 - mous
40	43	48.9	796	2	I49556	cadherin-11 - mous
41	43	48.9	796	2	A53584	OB-cadherin precur
42	43	48.9	601	2	A27020	DIF-induced presta
43	42.5	48.3	135	2	S63377	hypothetical prote
44	42	47.7	285	2	B71194	probable integrase
45	41	46.6	394	2	H64448	polyferredoxin - M

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 88; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CECNIKVKDVNDNFP 15
           |||||||||||||||
Db     251 CECNIKVKDVNDNFP 265
```

RESULT 2
IJHUG1
desmoglein 1 precursor - human
N;Alternate names: desmosomal glycoprotein I
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S16906; A39706; A61254; A61279; S16158
R;Buxton, R.S.
submitted to the EMBL Data Library, November 1990
A;Reference number: S16906
A;Accession: S16906
A;Molecule type: mRNA

A;Residues: 1-1049 <BUX>
A;Cross-references: EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
R;Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; R
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991
A;Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,
A;Reference number: A39706; MUID:91271279
A;Accession: A39706
A;Molecule type: mRNA
A;Residues: 24-1049 <WHE>
A;Cross-references: GB:X56654
R;Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
J. Cell Sci. 99, 809-821, 1991
A;Title: Structural analysis and expression of human desmoglein: a cadherin-like compone
A;Reference number: A61254; MUID:92121251
A;Accession: A61254
A;Molecule type: mRNA
A;Residues: 26-1049 <NIL>
R;Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee
Biochem. Soc. Trans. 19, 1060-1064, 1991
A;Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfami
A;Reference number: A61279; MUID:92175187
A;Accession: A61279
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-55 <WH3>
C;Genetics:
A;Gene: GDB:DSG1
A;Cross-references: GDB:126563; OMIM:125670
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prc
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1049/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-493/Domain: cadherin repeat homology <CR4>
F;509-530/Region: serine/threonine-rich
F;549-569/Domain: transmembrane #status predicted <TMM>
F;572-1049/Domain: intracellular #status predicted <INT>
F;840-869/Domain: desmoglein repeat <DG1>
F;870-899/Domain: desmoglein repeat <DG2>
F;900-927/Domain: desmoglein repeat <DG3>
F;928-956/Domain: desmoglein repeat <DG4>
F;969-1019/Region: glycine/serine-rich
F;110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.1%; Score 74; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 0.0004;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
|||||| ||||| |
Db 253 CECNIKILDVNDNIP 267

RESULT 3
IJBOG1
desmoglein 1 precursor - bovine
N;Alternate names: desmoglein BDGM
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S14603; A38872; A37785; S38721; A48173; S24412
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
submitted to the EMBL Data Library, March 1991
A;Description: Complete sequence of the desmoglein precursor and evidence for the existe
A;Reference number: S14603
A;Accession: S14603
A;Molecule type: mRNA
A;Residues: 1-1043 <KOC>
A;Cross-references: EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti
A;Reference number: A38872; MUID:92037656
A;Accession: A38872
A;Molecule type: mRNA
A;Residues: 1-87;968-1043 <KO2>
A;Cross-references: GB:S64268; GB:S64270
R;Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990
A;Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion
A;Reference number: A37785; MUID:91097553
A;Accession: A37785
A;Molecule type: mRNA
A;Residues: 44-123,'V',125-493 <GOO>
A;Cross-references: GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
R;Zimbelmann, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S38721
A;Accession: S38721
A;Molecule type: mRNA
A;Residues: 44-1043 <ZIM>
A;Cross-references: EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
R;Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.
Eur. J. Cell Biol. 53, 1-12, 1990
A;Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m
A;Reference number: A48173; MUID:91168965
A;Accession: A48173
A;Molecule type: mRNA
A;Residues: 44-1001,'AQPPSAT' <KO3>
A;Cross-references: GB:X57784
A;Note: this sequence has been revised in references A38872 and S38721
C;Genetics:
A;Gene: DSG1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1043/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-491/Domain: cadherin repeat homology <CR4>
F;549-574/Domain: transmembrane #status predicted <TMM>
F;575-1043/Domain: intracellular #status predicted <INT>
F;846-875/Domain: desmoglein repeat <DG1>
F;876-905/Domain: desmoglein repeat <DG2>
F;906-933/Domain: desmoglein repeat <DG3>
F;934-962/Domain: desmoglein repeat <DG4>
F;963-1012/Region: glycine/serine-rich
F;110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.1%; Score 74; DB 1; Length 1043;
Best Local Similarity 80.0%; Pred. No. 0.0004;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CECNIKVKDVNDNFP 15
|||||| ||||| |
Db 253 CECNIKILDVNDNIP 267

RESULT 4
S55396
LI-cadherin - human
C;Species: Homo sapiens (man)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Sep-1997
C;Accession: S55396
R;Boettinger, A.; Kreft, B.; Fieger, C.; Dlouhy, B.; Berndorff, D.; Goessner, R.; Tau
submitted to the EMBL Data Library, December 1994

```
A;Description: Molecular cloning of human LI-cadherin:evidence for a novel type of cadhe
A;Reference number: S55396
A;Accession: S55396
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-832 <BOE>
A;Cross-references: EMBL:X83228; NID:g854174; PID:g854175
C;Superfamily: cadherin repeat homology
F;455-566/Domain: cadherin repeat homology <CR3>

  Query Match           59.1%;  Score 52;  DB 2;  Length 832;
  Best Local Similarity  64.3%;  Pred. No. 1.1;
  Matches    9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ECNIKVKDVNDNFP 15
          | ::|||||:||| |
Db    325 EIHVKVKDINDNPP 338


RESULT   5
T15276
hypothetical protein R10F2.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15276
R;Du, Z.; Gattung, S.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid R10F2.
A;Reference number: Z18320
A;Accession: T15276
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2163 <DUZ>
A;Cross-references: EMBL:AF003388; NID:g2088850; PID:g2088852; PIDN:AAB54266.1; GSPDB:GN
A;Experimental source: strain Bristol N2; clone R10F2
C;Genetics:
A;Gene: CESP:R10F2.1
A;Map position: 3
A;Introns: 34/1; 114/3; 735/3; 1552/2; 1800/3; 1865/3; 1911/3; 2026/3; 2108/2; 2150/3

  Query Match           58.0%;  Score 51;  DB 2;  Length 2163;
  Best Local Similarity  53.3%;  Pred. No. 4.1;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |:|:: | | ||| |
Db   1629 CKCHVIVLDENDNLP 1643


RESULT   6
IJBODC
desmocollin 2a precursor - bovine (fragment)
N;Alternate names: epithelial type 2 desmocollin subform I
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41799
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky, R.; Franke, W.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 353-357, 1992
A;Title: Complexity and expression patterns of the desmosomal cadherins.
A;Reference number: A41799; MUID:92108053
A;Accession: A41799
A;Molecule type: mRNA
A;Residues: 1-863 <KOC>
A;Cross-references: GB:M81190; NID:g163757; PIDN:AAA30782.1; PID:g163758
A;Experimental source: muzzle
A;Note: sequence extracted from NCBI backbone
A;Note: 264-Gln and 333-Gln were also found
C;Genetics:
A;Gene: DSC2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprot
F;1-89/Domain: propeptide #status predicted <PRO>
F;90-863/Product: desmocollin 2a #status predicted <MAT>
F;90-645/Domain: extracellular #status predicted <EXT>
F;92-197/Domain: cadherin repeat homology <CR1>
F;200-309/Domain: cadherin repeat homology <CR2>
F;312-423/Domain: cadherin repeat homology <CR3>
F;426-526/Domain: cadherin repeat homology <CR4>
F;527-634/Domain: cadherin repeat homology <CR5>
F;646-672/Domain: transmembrane #status predicted <TMM>
F;673-863/Domain: intracellular #status predicted <INT>
F;120,346,495,579/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;826/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match           55.7%;  Score 49;  DB 1;  Length 863;
  Best Local Similarity  61.5%;  Pred. No. 3.5;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 CNIKVKDVNDNFP 15
          | | ::||||| |
Db    295 CIINIEDVNDNLP 307


RESULT   7
IJBODD
desmocollin 2b precursor - bovine (fragment)
N;Alternate names: epithelial type 2 desmocollin subform II
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: B41799
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky, R.; Franke, W.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 353-357, 1992
A;Title: Complexity and expression patterns of the desmosomal cadherins.
A;Reference number: A41799; MUID:92108053
A;Accession: B41799
A;Molecule type: mRNA
A;Residues: 1-809 <KOC>
A;Cross-references: GB:M81190; NID:g163757; PIDN:AAA30783.1; PID:g163759
A;Experimental source: muzzle
A;Note: sequence extracted from NCBI backbone
A;Note: 264-Gln and 333-Gln were also found
C;Genetics:
A;Gene: DSC2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycop
F;1-89/Domain: propeptide #status predicted <PRO>
F;90-809/Product: desmocollin 2b #status predicted <MAT>
F;90-645/Domain: extracellular #status predicted <EXT>
F;92-197/Domain: cadherin repeat homology <CR1>
F;200-309/Domain: cadherin repeat homology <CR2>
F;312-423/Domain: cadherin repeat homology <CR3>
F;426-526/Domain: cadherin repeat homology <CR4>
F;527-634/Domain: cadherin repeat homology <CR5>
F;646-672/Domain: transmembrane #status predicted <TMM>
F;673-809/Domain: intracellular #status predicted <INT>
F;120,346,495,579/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match           55.7%;  Score 49;  DB 1;  Length 809;
  Best Local Similarity  61.5%;  Pred. No. 3.3;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 CNIKVKDVNDNFP 15
          | | ::||||| |
Db    295 CIINIEDVNDNLP 307


RESULT   8
IJFFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C;Accession: A41087; B41087
```

```
R;Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A;Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A;Reference number: A41087; MUID:92069752
A;Accession: A41087
A;Molecule type: mRNA
A;Residues: 143-485;1279-5147 <MAH>
A;Cross-references: GB:M80537
A;Accession: B41087
A;Molecule type: DNA
A;Residues: 1-142;487-1278 <MA2>
A;Cross-references: GB:M80537
A;Note: 1229-Gly and 1233-Ser were also found
C;Genetics:
A;Gene: fat
A;Cross-references: FlyBase:FBgn0001075
C;Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C;Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
F;36-4583/Domain: extracellular #status predicted <EXT>
F;51-156/Domain: cadherin repeat homology <CR1>
F;159-270/Domain: cadherin repeat homology <CR2>
F;271-382/Domain: cadherin repeat homology <CR3>
F;390-494/Domain: cadherin repeat homology <CR4>
F;497-599/Domain: cadherin repeat homology <CR5>
F;602-708/Domain: cadherin repeat homology <CR6>
F;718-822/Domain: cadherin repeat homology <CR7>
F;831-942/Domain: cadherin repeat homology <CR8>
F;948-1049/Domain: cadherin repeat homology <CR9>
F;1052-1153/Domain: cadherin repeat homology <C10>
F;1156-1278/Domain: cadherin repeat homology <C11>
F;1281-1384/Domain: cadherin repeat homology <C12>
F;1387-1489/Domain: cadherin repeat homology <C13>
F;1492-1601/Domain: cadherin repeat homology <C14>
F;1607-1713/Domain: cadherin repeat homology <C15>
F;1717-1823/Domain: cadherin repeat homology <C16>
F;1826-1922/Domain: cadherin repeat homology <C17>
F;1925-2027/Domain: cadherin repeat homology <C18>
F;2028-2167/Domain: cadherin repeat homology <C19>
F;2169-2278/Domain: cadherin repeat homology <C20>
F;2281-2384/Domain: cadherin repeat homology <C99>
F;2387-2491/Domain: cadherin repeat homology <C21>
F;2494-2596/Domain: cadherin repeat homology <C22>
F;2599-2703/Domain: cadherin repeat homology <C23>
F;2707-2810/Domain: cadherin repeat homology <C24>
F;2813-2913/Domain: cadherin repeat homology <C25>
F;2915-3013/Domain: cadherin repeat homology <C26>
F;3014-3124/Domain: cadherin repeat homology <C27>
F;3127-3229/Domain: cadherin repeat homology <C28>
F;3232-3334/Domain: cadherin repeat homology <C29>
F;3337-3439/Domain: cadherin repeat homology <C30>
F;3442-3545/Domain: cadherin repeat homology <C31>
F;3548-3651/Domain: cadherin repeat homology <C32>
F;3654-3756/Domain: cadherin repeat homology <C33>
F;3954-4010/Domain: EGF homology <EG1>
F;4017-4048/Domain: EGF homology <EG2>
F;4056-4089/Domain: EGF homology <EG3>
F;4096-4127/Domain: EGF homology <EG4>
F;4584-4609/Domain: transmembrane #status predicted <TMM>
F;4610-5147/Domain: intracellular #status predicted <INT>


  Query Match             54.5%;  Score 48;  DB 1;  Length 5147;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches   10;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 ECNIKVKDVNDNFP 15
          |  ||| ||||| |
Db    141 EVRIKVLDVNDNSP 154


RESULT   9
IJHUC5
cadherin 5 precursor - human
N;Alternate names: 7B4 antigen; cadherin, endothelial-specific; VE-cadherin
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C;Accession: S49893; S24305; A43418
R;Breviario, F.; Caveda, L.; Corada, M.; Martin-Padura, I.; Golay, J.; Introna, M.; L
submitted to the EMBL Data Library, June 1994
A;Description: Molecular and functional properties of VE-cadherin (7B4/cadherin-5) a
A;Reference number: S49893
A;Accession: S49893
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-784 <BRE>
A;Cross-references: EMBL:X79981; NID:g599833; PIDN:CAA56306.1; PID:g599834
R;Suzuki, S.; Sano, K.; Tanihara, H.
Cell Regul. 2, 261-270, 1991
A;Title: Diversity of the cadherin family: evidence for eight new cadherins in nervou
A;Reference number: S24305; MUID:91283540
A;Accession: S24305
A;Molecule type: mRNA
A;Residues: 5-516,'I',518-784 <SUZ>
A;Cross-references: EMBL:X59796; NID:g639976; PIDN:CAA42468.1; PID:g29593
R;Lampugnani, M.G.; Resnati, M.; Raiteri, M.; Pigott, R.; Pisacane, A.; Houen, G.; Ru
J. Cell Biol. 118, 1511-1522, 1992
A;Title: A novel endothelial-specific membrane protein is a marker of cell-cell conta
A;Reference number: A43418; MUID:92394977
A;Accession: A43418
A;Molecule type: protein
A;Residues: 48-60,'X',62,'X',64;108-116,'X',118-123;237-238,'X',240,'X',242-252,'X',2
A;Experimental source: cultured endothelial cells
A;Note: sequence extracted from NCBI backbone (NCBIP:113040, NCBIP:113045, NCBIP:1130
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though
C;Genetics:
A;Gene: GDB:CDH5
A;Cross-references: GDB:134230; OMIM:601120
A;Map position: 16q22.1-16q22.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-47/Domain: propeptide #status predicted <PRO>
F;48-784/Product: cadherin 5 #status predicted <MAT>
F;48-593/Domain: extracellular #status predicted <EXT>
F;50-151/Domain: cadherin repeat homology <CR1>
F;154-258/Domain: cadherin repeat homology <CR2>
F;261-372/Domain: cadherin repeat homology <CR3>
F;375-479/Domain: cadherin repeat homology <CR4>
F;481-587/Domain: cadherin repeat homology <CR5>
F;594-620/Domain: transmembrane #status predicted <TMM>
F;621-784/Domain: intracellular #status predicted <INT>
F;736-753/Region: serine-rich
F;61,112,157,362,442,523,535/Binding site: carbohydrate (Asn) (covalent) #status pred


  Query Match             54.5%;  Score 48;  DB 1;  Length 784;
  Best Local Similarity   81.8%;  Pred. No. 4.6;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 IKVKDVNDNFP 15
          ||| ||||| |
Db    139 IKVHDVNDNWP 149


RESULT  10
I45858
desmocollin - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999
C;Accession: I45858
R;Yue, K.K.; Holton, J.L.; Clarke, J.P.; Hyam, J.L.; Hashimoto, T.; Chidgey, M.A.; Ga
J. Cell Sci. 108, 2163-2173, 1995
A;Title: Characterisation of a desmocollin isoform (bovine DSC3) exclusively expresse
A;Reference number: I45858; MUID:95403557
```

```
A;Accession: I45858
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-896 <YUE>
A;Cross-references: GB:L33774; NID:g914820; PIDN:AAC41625.1; PID:g914821
C;Genetics:
A;Gene: Dsc3
A;Introns: 831/3
C;Superfamily: cadherin; cadherin repeat homology

  Query Match             53.4%;  Score 47;  DB 2;  Length 896;
  Best Local Similarity   69.2%;  Pred. No. 7.6;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CNIKVKDVNDNFP 15
           | | ||| ||| |
Db     340 CIITVKDSNDNLP 352


RESULT  11
S44887
ZK112.7 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C;Accession: S44887
R;Du, Z.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid ZK112.
A;Reference number: S44613
A;Accession: S44887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3343 <DUZ>
A;Cross-references: EMBL:L14324; NID:g289740; PID:g289742
C;Genetics:
A;Introns: 39/1; 74/3; 122/3; 209/2; 287/3; 333/2; 489/1; 611/2; 714/1; 939/3; 1002/2; 1
C;Keywords: cytoskeleton; transmembrane protein

  Query Match             53.4%;  Score 47;  DB 2;  Length 3343;
  Best Local Similarity   53.3%;  Pred. No. 28;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |  || : |:||| |
Db     213 CHLNILILDINDNPP 227


RESULT  12
S38673
desmoglein 2 - human
N;Alternate names: desmoglein HDGC
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C;Accession: S38673; B38872
R;Zimbelmann, R.
submitted to the EMBL Data Library, September 1993
A;Reference number: S38673
A;Accession: S38673
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1117 <ZIM>
A;Cross-references: EMBL:Z26317; NID:g416177; PIDN:CAA81226.1; PID:g416178
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide
A;Reference number: A38872; MUID:92037656
A;Accession: B38872
A;Molecule type: mRNA
A;Residues: 777-1117 <KOC>
A;Cross-references: GB:S64273
C;Genetics:
A;Gene: GDB:DSG2
A;Cross-references: GDB:128808; OMIM:125671
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane prote
F;51-158/Domain: cadherin repeat homology <CR1>
F;161-271/Domain: cadherin repeat homology <CR2>

  Query Match             52.3%;  Score 46;  DB 2;  Length 1117;
  Best Local Similarity   64.3%;  Pred. No. 14;
  Matches    9;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ECNIKVKDVNDNFP 15
           |  ||| |:||| |
Db     143 ELRIKVLDINDNEP 156


RESULT  13
D38992
cadherin 8 - human
C;Species: Homo sapiens (man)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
C;Accession: D38992
R;Suzuki, S.; Sano, K.; Tanihara, H.
Cell Regul. 2, 261-270, 1991
A;Title: Diversity of the cadherin family: evidence for eight new cadherins in nervou
A;Reference number: S24305; MUID:91283540
A;Accession: D38992
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-793 <SUZ>
A;Cross-references: GB:L34060; NID:g506411; PIDN:AAA35628.1; PID:g506412
C;Genetics:
A;Gene: GDB:CDH8
A;Cross-references: GDB:5822911
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication

  Query Match             52.3%;  Score 46;  DB 2;  Length 793;
  Best Local Similarity   72.7%;  Pred. No. 9.7;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     148 IKVQDINDNAP 158


RESULT  14
G02678
cadherin-14 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C;Accession: G02678
R;Shibata, T.; Shimoyama, Y.; Gotoh, M.; Hirohashi, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: H01584
A;Accession: G02678
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-790 <SHI>
A;Cross-references: EMBL:U59325; NID:g1389852; PIDN:AAB02933.1; PID:g1389853
C;Superfamily: cadherin; cadherin repeat homology

  Query Match             52.3%;  Score 46;  DB 2;  Length 790;
  Best Local Similarity   72.7%;  Pred. No. 9.7;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     147 IKVQDINDNAP 157
```

RESULT 15
T09055
protocadherin 68 - human
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T09055
R;Jin, P.; Xu, H.; Israel, D.
submitted tq the EMBL Data Library, October 1997
A;Reference number: Z16540
A;Accession: T09055
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-889 <JIN>
A;Cross-references: EMBL:AF029343; NID:g2599501; PID:g2599502
C;Genetics:
A;Gene: PCH68

Query Match 52.3%; Score 46; DB 2; Length 889;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 CECNIKVKDVNDNFP 15
           |   ::::|:||| |
Db     116 CMIKVEIQDINDNAP 130
```

Search completed: January 7, 2000, 08:52:44
Job time: 2274 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:39 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-5
Perfect score: 88
Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	88	100.0	999	1	DSG3_HUMAN	P32926	homo sapien
2	74	84.1	1043	1	DSG1_BOVIN	Q03763	bos taurus
3	74	84.1	1049	1	DSG1_HUMAN	Q02413	homo sapien
4	49	55.7	863	1	DSC2_BOVIN	P33545	bos taurus
5	49	55.7	902	1	DSC2_MOUSE	P55292	mus musculu
6	48	54.5	784	1	CAD5_HUMAN	P33151	homo sapien
7	48	54.5	783	1	CAD5_MOUSE	P55284	mus musculu
8	48	54.5	5147	1	FAT_DROME	P33450	drosophila
9	47	53.4	782	1	CAD5_PIG	O02840	sus scrofa
10	47	53.4	896	1	DSC3_BOVIN	Q28060	bos taurus
11	47	53.4	3343	1	YOG7_CAEEL	P34616	caenorhabdi
12	46	52.3	793	1	CAD8_HUMAN	P55286	homo sapien
13	46	52.3	799	1	CAD8_MOUSE	P97291	mus musculu
14	46	52.3	790	1	CADE_HUMAN	Q13634	homo sapien
15	46	52.3	1117	1	DSG2_HUMAN	Q14126	homo sapien
16	45	51.1	794	1	CADC_HUMAN	P55289	homo sapien
17	45	51.1	827	1	CADL_RAT	P55281	rattus norv
18	45	51.1	901	1	DSC2_HUMAN	Q02487	homo sapien
19	45	51.1	805	1	IF2_AQUAE	O67825	aquifex aeo
20	44	50.0	884	1	CAD1_MOUSE	P09803	mus musculu
21	44	50.0	790	1	CAD6_HUMAN	P55285	homo sapien
22	44	50.0	790	1	CAD6_MOUSE	P97326	mus musculu
23	44	50.0	789	1	CAD6_RAT	P55280	rattus norv
24	44	50.0	3051	1	YNX3_CAEEL	P34576	caenorhabdi
25	43	48.9	796	1	CADB_HUMAN	P55287	homo sapien
26	43	48.9	796	1	CADB_MOUSE	P55288	mus musculu
27	42.5	48.3	135	1	YN8Q_YEAST	P53738	saccharomyc
28	42	47.7	789	1	CADA_CHICK	P79995	gallus gall
29	41	46.6	872	1	CAD1_XENLA	P30944	xenopus lae
30	41	46.6	171	1	PCR1_SCHPO	Q09926	schizosacch
31	40	45.5	732	1	CADB_CHICK	P33145	gallus gall
32	40	45.5	712	1	CADD_CHICK	P33150	gallus gall
33	40	45.5	527	1	YAD7_SCHPO	Q09833	schizosacch
34	39.5	44.9	1986	1	WA_EMENI	Q03149	emericella
35	39	44.3	887	1	CAD1_CHICK	P08641	gallus gall
36	39	44.3	426	1	CISY_HELPY	P56062	helicobacte
37	39	44.3	894	1	DSC1_HUMAN	Q08554	homo sapien
38	39	44.3	896	1	DSC3_HUMAN	Q14574	homo sapien
39	39	44.3	294	1	DSC3_MOUSE	P55850	mus musculu
40	39	44.3	780	1	OBP_HSV6U	P52378	herpes simp
41	39	44.3	780	1	OBP_HSV6Z	P52452	herpes simp
42	39	44.3	82	1	PL05_PLETR	P36983	plectreurys
43	39	44.3	46	1	PL08_PLETR	P36984	plectreurys
44	39	44.3	79	1	PL11_PLETR	P36985	plectreurys
45	39	44.3	45	1	PL12_PLETR	P36986	plectreurys

ALIGNMENTS

RESULT 1
DSG3_HUMAN
ID DSG3_HUMAN STANDARD; PRT; 999 AA.
AC P32926;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92069753.
RA AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion.";
RL Cell 67:869-877(1991).
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC ---

CC ---
DR EMBL; M76482; AAA60230.1; -.
DR PIR; A41088; IJHUG3.
DR HSSP; P09803; 1EDH.
DR MIM; 169615; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.

FT	SIGNAL	1	23 POTENTIAL.
FT	PROPEP	24	49 POTENTIAL.
FT	CHAIN	50	999 DESMOGLEIN 3.
FT	DOMAIN	50	615 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	616	640 POTENTIAL.
FT	DOMAIN	641	999 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50	158 CADHERIN 1.
FT	REPEAT	159	268 CADHERIN 2.
FT	REPEAT	269	383 CADHERIN 3.
FT	REPEAT	386	499 CADHERIN 4.
FT	REPEAT	910	935 DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966 DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110	110 POTENTIAL.

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match             100.0%;  Score 88;  DB 1;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |||||||||||||||
Db    251 CECNIKVKDVNDNFP 265


RESULT   2
DSG1_BOVIN
ID   DSG1_BOVIN     STANDARD;      PRT;  1043 AA.
AC   Q03763;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
GN   DSG1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANKE W.W.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 44-1043 FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 91168965.
RA   KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA   ZIMBELMANN R., FRANKE W.W.;
RT   "Identification of desmoglein, a constitutive desmosomal
RT   glycoprotein, as a member of the cadherin family of cell adhesion
RT   molecules.";
RL   Eur. J. Cell Biol. 53:1-12(1990).
RN   [3]
RP   REVISIONS, AND SEQUENCE OF 101-123.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA   FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
RN   [4]
RP   SEQUENCE OF 44-493 FROM N.A.
RX   MEDLINE; 91097553.
RA   GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT   "Desmoglein shows extensive homology to the cadherin family of cell
RT   adhesion molecules.";
RL   Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X58466; CAA41380.1; -.
DR   EMBL; X57784; CAA40930.1; -.
DR   EMBL; M58165; AAA62709.1; -.
DR   PIR; S14603; IJBOG1.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1043       DESMOGLEIN 1.
FT   DOMAIN       50    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    549    573       POTENTIAL.
FT   DOMAIN      574   1043       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    498       CADHERIN 4.
FT   REPEAT      819    845       DESMOGLEIN REPEAT 1.
FT   REPEAT      846    875       DESMOGLEIN REPEAT 2.
FT   REPEAT      876    905       DESMOGLEIN REPEAT 3.
FT   REPEAT      906    933       DESMOGLEIN REPEAT 4.
FT   REPEAT      934    962       DESMOGLEIN REPEAT 5.
FT   DOMAIN      963   1012       GLY/SER-RICH.
FT   CARBOHYD    110    110
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    496    496       POTENTIAL.
FT   CONFLICT    124    124       I -> V (IN REF. 4).
SQ   SEQUENCE   1043 AA;  112243 MW;  13898584 CRC32;


  Query Match             84.1%;  Score 74;  DB 1;  Length 1043;
  Best Local Similarity   80.0%;  Pred. No. 0.00016;
  Matches   12;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          ||||||: ||||| |
Db    253 CECNIKILDVNDNIP 267


RESULT   3
DSG1_HUMAN
ID   DSG1_HUMAN     STANDARD;      PRT;  1049 AA.
AC   Q02413;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN   DSG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 91271279.
RA   WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOTIS P., POYNTER D.,
RA   ARNEMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA   BUXTON R.S., MAGEE A.I.;
RT   "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT   junctions, is related to the cadherin family of cell adhesion
RT   molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
```

```
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56654; CAA39976.1; -.
DR   PIR; S16906; IJHUG1.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125670; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1049       DESMOGLEIN 1.
FT   DOMAIN       50    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    570       POTENTIAL.
FT   DOMAIN      571   1049       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    497       CADHERIN 4.
FT   REPEAT      813    839       DESMOGLEIN REPEAT 1.
FT   REPEAT      840    869       DESMOGLEIN REPEAT 2.
FT   REPEAT      870    899       DESMOGLEIN REPEAT 3.
FT   REPEAT      900    927       DESMOGLEIN REPEAT 4.
FT   REPEAT      928    956       DESMOGLEIN REPEAT 5.
FT   DOMAIN      969   1019       GLY/SER-RICH.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    180    180       POTENTIAL.
SQ   SEQUENCE   1049 AA;  113715 MW;  FDD79961 CRC32;


  Query Match             84.1%;  Score 74;  DB 1;  Length 1049;
  Best Local Similarity   80.0%;  Pred. No. 0.00016;
  Matches   12;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |||||| ||||| |
Db     253 CECNIKILDVNDNIP 267


RESULT   4
DSC2_BOVIN
ID   DSC2_BOVIN     STANDARD;      PRT;   863 AA.
AC   P33545;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN)
DE   (FRAGMENT).
GN   DSC2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 92108053.
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., TROYANOVSKY R.,
RA   FRANKE W.W.;
RT   "Complexity and expression patterns of the desmosomal cadherins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:353-357(1992).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: FORMS 2A (SHOWN HERE) AND 2B ARE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- TISSUE SPECIFICITY: ESOPHAGUS AND RUMEN. WEAKLY IN EPITHELIA AND
CC       CARDIAC MUSCLE.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81190; AAA30782.1; -.
DR   EMBL; M81190; AAA30783.1; -.
DR   PIR; A41799; IJBODC.
DR   PIR; B41799; IJBODD.
DR   HSSP; P09803; 1SUH.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Alternative splicing; Cytoskeleton; Calcium-binding.
FT   NON_TER       1      1
FT   PROPEP        1     89       POTENTIAL.
FT   CHAIN        90    863       DESMOCOLLIN 2A/2B.
FT   DOMAIN       90    644       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    645    665       POTENTIAL.
FT   DOMAIN      666    863       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       90    197       CADHERIN 1.
FT   REPEAT      198    309       CADHERIN 2.
FT   REPEAT      310    423       CADHERIN 3.
FT   REPEAT      424    528       CADHERIN 4.
FT   REPEAT      529    644       CADHERIN 5.
FT   CARBOHYD    120    120       POTENTIAL.
FT   CARBOHYD    346    346       POTENTIAL.
FT   CARBOHYD    495    495       POTENTIAL.
FT   CARBOHYD    579    579       POTENTIAL.
FT   VARIANT     264    264       K -> Q.
FT   VARIANT     333    333       R -> Q.
FT   VARSPLIC    799    809       KVQQCDQDNTH -> ESIRGHTLVKN (IN FORM 2B).
FT   VARSPLIC    810    863       MISSING (IN FORM 2B).
SQ   SEQUENCE   863 AA;  95874 MW;  2173F06E CRC32;


  Query Match             55.7%;  Score 49;  DB 1;  Length 863;
  Best Local Similarity   61.5%;  Pred. No. 1.6;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CNIKVKDVNDNFP 15
           | | ::||||| |
Db     295 CIINIEDVNDNLP 307


RESULT   5
DSC2_MOUSE
ID   DSC2_MOUSE     STANDARD;      PRT;   902 AA.
AC   P55292; Q64734;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN).
GN   DSC2 OR DSC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95227276.
RA   LORIMER J.E., HALL L.S., CLARKE J.P., COLLINS J.E., FLEMING T.P.,
RA   GARROD D.R.;
RT   "Cloning, sequence analysis and expression pattern of mouse
RT   desmocollin 2 (DSC2), a cadherin-like adhesion molecule.";
RL   Mol. Membr. Biol. 11:229-236(1994).
RN   [2]
RP   SEQUENCE OF 344-637 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=EMBRYO;
RX   MEDLINE; 95048328.
RA   BUXTON R.S., WHEELER G.N., PIDSLEY S.C., MARSDEN M.D., ADAMS M.J.,
RA   JENKINS N.A., GILBERT D.J., COPELAND N.G.;
RT   "Mouse desmocollin (Dsc3) and desmoglein (Dsg1) genes are closely
RT   linked in the proximal region of chromosome 18.";
RL   Genomics 21:510-516(1994).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: FORMS 2A AND 2B ARE PRODUCED BY ALTERNATIVE
CC       SPLICING OF THE SAME GENE.
CC   -!- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L33779; AAA79177.1; -.
DR   EMBL; L33779; AAA79176.1; -.
DR   EMBL; X73885; CAA52089.1; -.
DR   HSSP; P09803; 1EDH.
DR   MGD; MGI:103221; DSC2.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Cytoskeleton; Calcium-binding.
FT   SIGNAL        1     27       POTENTIAL.
FT   PROPEP       28    135       POTENTIAL.
FT   CHAIN       136    902       DESMOCOLLIN 2A/2B.
FT   DOMAIN      136    694       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    695    715       POTENTIAL.
FT   DOMAIN      716    902       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      136    243       CADHERIN 1.
FT   REPEAT      244    355       CADHERIN 2.
FT   REPEAT      356    471       CADHERIN 3.
FT   REPEAT      472    579       CADHERIN 4.
FT   REPEAT      580    694       CADHERIN 5.
FT   CARBOHYD    166    166       POTENTIAL.
FT   CARBOHYD    392    392       POTENTIAL.
FT   CARBOHYD    546    546       POTENTIAL.
FT   CARBOHYD    629    629       POTENTIAL.
FT   VARSPLIC    838    848       KVQFCHTDDNQ -> ETIRGHTLIKN (IN FORM 2B).
FT   VARSPLIC    849    902       MISSING (IN FORM 2B).
SQ   SEQUENCE   902 AA;  99961 MW;  37FE753B CRC32;


  Query Match             55.7%;  Score 49;  DB 1;  Length 902;
  Best Local Similarity   57.1%;  Pred. No. 1.7;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ECNIKVKDVNDNFP 15
           :| | ::|||||  |
Db     340 KCIITIEDVNDNLP 353


RESULT   6
CAD5_HUMAN
ID   CAD5_HUMAN     STANDARD;      PRT;   784 AA.
AC   P33151;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
DE   (7B4 ANTIGEN) (CD144 ANTIGEN).
GN   CDH5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ENDOTHELIAL CELLS;
RX   MEDLINE; 95353875.
RA   BREVIARIO F., CAVEDA L., CORADA M., MARTIN-PADURA I., NAVARRO P.,
RA   GOLAY J., INTRONA M., GULINO D., LAMPUGNANI M.G., DEJANA E.;
RT   "Functional properties of human vascular endothelial cadherin
RT   (7B4/cadherin-5), an endothelium-specific cadherin.";
RL   Arterioscler. Thromb. Vasc. Biol. 15:1229-1239(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   ALI J., MULLER W.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 5-784 FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 91283540.
RA   SUZUKI S., SANO K., TANIHARA H.;
RT   "Diversity of the cadherin family: evidence for eight new cadherins
RT   in nervous tissue.";
RL   Cell Regul. 2:261-270(1991).
RN   [4]
RP   PARTIAL SEQUENCE.
RC   TISSUE=ENDOTHELIAL CELLS;
RX   MEDLINE; 92394977.
RA   LAMPUGNANI M.G., RESNATI M., RAITERI M., PIGOTT R., PISACANE A.,
RA   HOUEN G., RUCO L.P., DEJANA E.;
RT   "A novel endothelial-specific membrane protein is a marker of
RT   cell-cell contacts.";
RL   J. Cell Biol. 118:1511-1522(1992).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC       IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC       COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
CC       ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
CC       BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES.
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL TISSUES AND BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD144 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd144.htm".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X79981; CAA56306.1; -.
DR   EMBL; U84722; AAB41796.1; -.
DR   EMBL; X59796; CAA42468.1; -.
DR   PIR; S24305; IJHUC5.
```

```
DR   HSSP; P09803; 1EDH.
DR   MIM; 601120; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   PROPEP      26     47       POTENTIAL.
FT   CHAIN       48    784       VASCULAR ENDOTHELIAL-CADHERIN.
FT   DOMAIN      48    593       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   594    620       POTENTIAL.
FT   DOMAIN     621    784       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      48    151       CADHERIN 1.
FT   REPEAT     152    258       CADHERIN 2.
FT   REPEAT     259    372       CADHERIN 3.
FT   REPEAT     373    477       CADHERIN 4.
FT   REPEAT     478    593       CADHERIN 5.
FT   DOMAIN     736    753       SER-RICH.
FT   CARBOHYD    61     61       POTENTIAL.
FT   CARBOHYD   112    112       POTENTIAL.
FT   CARBOHYD   157    157       POTENTIAL.
FT   CARBOHYD   362    362       POTENTIAL.
FT   CARBOHYD   442    442       POTENTIAL.
FT   CARBOHYD   523    523       POTENTIAL.
FT   CARBOHYD   535    535       POTENTIAL.
FT   CONFLICT   517    517       I -> T (IN REF. 1).
SQ   SEQUENCE   784 AA;  87528 MW;  C2C5CD71 CRC32;
```

```
  Query Match           54.5%;  Score 48;  DB 1;  Length 784;
  Best Local Similarity  81.8%;  Pred. No. 2.1;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           ||| ||||| :|
Db     139 IKVHDVNDNWP 149
```

```
RESULT   7
CAD5_MOUSE
ID   CAD5_MOUSE     STANDARD;      PRT;   783 AA.
AC   P55284;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).
GN   CDH5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN CAPILLARY;
RX   MEDLINE; 96141083.
RA   BREIER G., BREVIARIO F., CAVEDA L., BERTHIER R., SCHNUERCH H.,
RA   GOTSCH U., VESTWEBER D., RISAU W., DEJANA E.;
RT   "Molecular cloning and expression of murine vascular endothelial-
RT   cadherin in early stage development of cardiovascular system.";
RL   Blood 87:630-641(1996).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC       IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC       COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. NERVOUS
CC       SYSTEM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83930; CAA58782.1; -.
DR   HSSP; P09803; 1EDH.
DR   MGD; MGI:105057; CDH5.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   PROPEP      25     45       POTENTIAL.
FT   CHAIN       46    783       VASCULAR ENDOTHELIAL-CADHERIN.
FT   DOMAIN      46    592       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   593    619       POTENTIAL.
FT   DOMAIN     620    783       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      46    148       CADHERIN 1.
FT   REPEAT     149    255       CADHERIN 2.
FT   REPEAT     256    370       CADHERIN 3.
FT   REPEAT     371    475       CADHERIN 4.
FT   REPEAT     476    592       CADHERIN 5.
FT   DOMAIN     737    752       SER-RICH.
FT   CARBOHYD    59     59       POTENTIAL.
FT   CARBOHYD   154    154       POTENTIAL.
FT   CARBOHYD   440    440       POTENTIAL.
FT   CARBOHYD   522    522       POTENTIAL.
FT   CARBOHYD   534    534       POTENTIAL.
SQ   SEQUENCE   783 AA;  87847 MW;  D0B71215 CRC32;
```

```
  Query Match           54.5%;  Score 48;  DB 1;  Length 783;
  Best Local Similarity  63.6%;  Pred. No. 2.1;
  Matches    7;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |:::|:|||||
Db     243 IRLEDINDNFP 253
```

```
RESULT   8
FAT_DROME
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN   FT.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92069752.
RA   MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA   GOODMAN C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   Cell 67:853-868(1991).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC       TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC       DIFFERENTIATION AND MORPHOGENESIS, AND DEAT DURING THE PUPAL
CC       STAGE.
CC   -!- MISCELLANEOUS: SIMILARITY; CONTAINS 37 CADHERINS-TYPE REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
```

```
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80537; AAA28530.1; -.
DR   PIR; A41087; IJFFTM.
DR   HSSP; P00740; 1IXA.
DR   FLYBASE; FBgn0001075; ft.
DR   PFAM; PF00008; EGF; 4.
DR   PFAM; PF00028; cadherin; 34.
DR   PFAM; PF00054; laminin_G; 2.
DR   PROSITE; PS00232; CADHERIN; 22.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat; EGF-like domain.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN       36   4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4584   4609       POTENTIAL.
FT   DOMAIN     4610   5147       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       36    156       CADHERIN 1.
FT   REPEAT      157    270       CADHERIN 2.
FT   REPEAT      271    382       CADHERIN 3.
FT   REPEAT      383    494       CADHERIN 4.
FT   REPEAT      495    599       CADHERIN 5.
FT   REPEAT      600    708       CADHERIN 6.
FT   REPEAT      709    820       CADHERIN 7.
FT   REPEAT      821    942       CADHERIN 8.
FT   REPEAT      943   1049       CADHERIN 9.
FT   REPEAT     1050   1153       CADHERIN 10.
FT   REPEAT     1154   1278       CADHERIN 11.
FT   REPEAT     1279   1384       CADHERIN 12.
FT   REPEAT     1385   1489       CADHERIN 13.
FT   REPEAT     1490   1601       CADHERIN 14.
FT   REPEAT     1602   1713       CADHERIN 15.
FT   REPEAT     1714   1823       CADHERIN 16.
FT   REPEAT     1824   1922       CADHERIN 17.
FT   REPEAT     1923   2027       CADHERIN 18.
FT   REPEAT     2028   2167       CADHERIN 19.
FT   REPEAT     2168   2278       CADHERIN 20.
FT   REPEAT     2279   2385       CADHERIN 21.
FT   REPEAT     2386   2491       CADHERIN 22.
FT   REPEAT     2492   2596       CADHERIN 23.
FT   REPEAT     2597   2703       CADHERIN 24.
FT   REPEAT     2704   2810       CADHERIN 25.
FT   REPEAT     2811   2913       CADHERIN 26.
FT   REPEAT     2914   3013       CADHERIN 27.
FT   REPEAT     3014   3124       CADHERIN 28.
FT   REPEAT     3125   3229       CADHERIN 29.
FT   REPEAT     3230   3334       CADHERIN 30.
FT   REPEAT     3335   3439       CADHERIN 31.
FT   REPEAT     3440   3545       CADHERIN 32.
FT   REPEAT     3546   3651       CADHERIN 33.
FT   REPEAT     3652   3756       CADHERIN 34.
FT   DOMAIN     3950   4011       EGF-LIKE 1.
FT   DOMAIN     4013   4049       EGF-LIKE 2.
FT   DOMAIN     4052   4090       EGF-LIKE 3.
FT   DOMAIN     4092   4128       EGF-LIKE 4.
FT   DOMAIN     4321   4362       EGF-LIKE 5.
FT   DISULFID   3954   3966       BY SIMILARITY.
FT   DISULFID   3960   3999       BY SIMILARITY.
FT   DISULFID   4001   4010       BY SIMILARITY.
FT   DISULFID   4017   4028       BY SIMILARITY.
FT   DISULFID   4022   4037       BY SIMILARITY.
FT   DISULFID   4039   4048       BY SIMILARITY.
FT   DISULFID   4056   4067       BY SIMILARITY.
FT   DISULFID   4061   4078       BY SIMILARITY.
FT   DISULFID   4080   4089       BY SIMILARITY.
FT   DISULFID   4096   4107       BY SIMILARITY.
FT   DISULFID   4101   4116       BY SIMILARITY.
FT   DISULFID   4118   4127       BY SIMILARITY.
FT   DISULFID   4325   4341       BY SIMILARITY.
FT   DISULFID   4334   4350       BY SIMILARITY.
FT   DISULFID   4352   4361       BY SIMILARITY.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    257    257       POTENTIAL.
FT   CARBOHYD    276    276       POTENTIAL.
FT   CARBOHYD    280    280       POTENTIAL.
FT   CARBOHYD    402    402       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    605    605       POTENTIAL.
FT   CARBOHYD    631    631       POTENTIAL.
FT   CARBOHYD   1155   1155       POTENTIAL.
FT   CARBOHYD   1367   1367       POTENTIAL.
FT   CARBOHYD   1458   1458       POTENTIAL.
FT   CARBOHYD   1751   1751       POTENTIAL.
FT   CARBOHYD   1831   1831       POTENTIAL.
FT   CARBOHYD   1880   1880       POTENTIAL.
FT   CARBOHYD   2080   2080       POTENTIAL.
FT   CARBOHYD   2171   2171       POTENTIAL.
FT   CARBOHYD   2247   2247       POTENTIAL.
FT   CARBOHYD   2290   2290       POTENTIAL.
FT   CARBOHYD   2437   2437       POTENTIAL.
FT   CARBOHYD   2581   2581       POTENTIAL.
FT   CARBOHYD   2799   2799       POTENTIAL.
FT   CARBOHYD   2920   2920       POTENTIAL.
FT   CARBOHYD   2946   2946       POTENTIAL.
FT   CARBOHYD   2967   2967       POTENTIAL.
FT   CARBOHYD   3167   3167       POTENTIAL.
FT   CARBOHYD   3303   3303       POTENTIAL.
FT   CARBOHYD   3386   3386       POTENTIAL.
FT   CARBOHYD   3389   3389       POTENTIAL.
FT   CARBOHYD   3525   3525       POTENTIAL.
FT   CARBOHYD   3852   3852       POTENTIAL.
FT   CARBOHYD   3865   3865       POTENTIAL.
FT   CARBOHYD   3905   3905       POTENTIAL.
FT   CARBOHYD   4306   4306       POTENTIAL.
FT   CARBOHYD   4414   4414       POTENTIAL.
FT   CARBOHYD   4471   4471       POTENTIAL.
FT   CARBOHYD   4487   4487       POTENTIAL.
FT   CARBOHYD   4539   4539       POTENTIAL.
FT   CARBOHYD   4550   4550       POTENTIAL.
FT   VARIANT    1229   1229       S -> G.
FT   VARIANT    1233   1233       G -> S.
SQ   SEQUENCE   5147 AA;  564868 MW;  1EF20E13 CRC32;


  Query Match             54.5%;  Score 48;  DB 1;  Length 5147;
  Best Local Similarity   71.4%;  Pred. No. 14;
  Matches   10;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ECNIKVKDVNDNFP 15
           |  ||| ||||| |
Db     141 EVRIKVLDVNDNSP 154


RESULT   9
CAD5_PIG
ID   CAD5_PIG       STANDARD;      PRT;   782 AA.
AC   O02840;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).
GN   CDH5.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   KILSHAW P.J.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC       IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC       COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
CC       ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
CC       BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13919; CAA74225.1; -.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   PROPEP      23     44       POTENTIAL.
FT   CHAIN       45    782       VASCULAR ENDOTHELIAL-CADHERIN.
FT   DOMAIN      45    592       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   593    619       POTENTIAL.
FT   DOMAIN     620    782       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      45    148       CADHERIN 1.
FT   REPEAT     149    255       CADHERIN 2.
FT   REPEAT     256    370       CADHERIN 3.
FT   REPEAT     371    475       CADHERIN 4.
FT   REPEAT     476    592       CADHERIN 5.
FT   DOMAIN     736    751       SER-RICH.
FT   CARBOHYD    58     58       POTENTIAL.
FT   CARBOHYD   154    154       POTENTIAL.
FT   CARBOHYD   360    360       POTENTIAL.
FT   CARBOHYD   440    440       POTENTIAL.
FT   CARBOHYD   522    522       POTENTIAL.
FT   CARBOHYD   534    534       POTENTIAL.
SQ   SEQUENCE   782 AA;  87546 MW;  007F70E0 CRC32;


  Query Match             53.4%;  Score 47;  DB 1;  Length 782;
  Best Local Similarity   72.7%;  Pred. No. 3.1;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           ||| |:|||:|
Db     136 IKVHDINDNWP 146


RESULT  10
DSC3_BOVIN
ID   DSC3_BOVIN     STANDARD;      PRT;   896 AA.
AC   Q28060; Q28061; Q28176;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOCOLLIN 3A/3B PRECURSOR.
GN   DSC3.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95403557.
RA   YUE K.K.M., HOLTON J.L., CLARKE J.P., HYAM J.L.M., HASHIMOTO T.,
RA   CHIDGEY M.A.J., GARROD D.R.;
RT   "Characterisation of a desmocollin isoform (bovine DSC3) exclusively
RT   expressed in lower layers of stratified epithelia.";
RL   J. Cell Sci. 108:2163-2173(1995).
RN   [2]
RP   SEQUENCE OF 686-814 FROM N.A.
RC   TISSUE=EPIDERMIS;
RX   MEDLINE; 94308280.
RA   LEGAN P.K., YUE K.K.M., CHIDGEY M.A.J., HOLTON J.L., WILKINSON R.W.,
RA   GARROD D.R.;
RT   "The bovine desmocollin family: a new gene and expression patterns
RT   reflecting epithelial cell proliferation and differentiation.";
RL   J. Cell Biol. 126:507-518(1994).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: FORMS 3A AND 3B ARE PRODUCED BY ALTERNATIVE
CC       SPLICING OF THE SAME GENE.
CC   -!- TISSUE SPECIFICITY: STRATIFIED EPITHELIA ONLY (EPIDERMIS, TONGUE,
CC       ESOPHAGUS AND RUMEN).
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L33774; AAC41625.1; -.
DR   EMBL; L33774; AAC41626.1; -.
DR   EMBL; X75783; CAA53427.1; -.
DR   HSSP; P09803; 1SUH.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Cytoskeleton; Calcium-binding.
FT   SIGNAL       1     26       POTENTIAL.
FT   PROPEP      27    134       POTENTIAL.
FT   CHAIN      135    896       DESMOCOLLIN 3A/3B.
FT   DOMAIN     135    690       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   691    711       POTENTIAL.
FT   DOMAIN     712    896       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     135    242       CADHERIN 1.
FT   REPEAT     243    354       CADHERIN 2.
FT   REPEAT     355    471       CADHERIN 3.
FT   REPEAT     472    579       CADHERIN 4.
FT   DOMAIN     580    690       CADHERIN 5.
FT   VARSPLIC   832    839       KLHLCNQD -> ESIRGHTG (IN FORM 3B).
FT   VARSPLIC   840    896       MISSING (IN FORM 3B).
FT   CARBOHYD   165    165       POTENTIAL.
FT   CARBOHYD   391    391       POTENTIAL.
FT   CARBOHYD   546    546       POTENTIAL.
FT   CARBOHYD   629    629       POTENTIAL.
FT   CONFLICT   686    687       VI -> EF (IN REF. 2).
SQ   SEQUENCE   896 AA;  99687 MW;  E5668408 CRC32;


  Query Match             53.4%;  Score 47;  DB 1;  Length 896;
  Best Local Similarity   69.2%;  Pred. No. 3.5;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CNIKVKDVNDNFP 15
           | | ||| ||| |
```

```
Db    340 CIITVKDSNDNLP 352


RESULT  11
YOG7_CAEEL
ID   YOG7_CAEEL     STANDARD;      PRT;  3343 AA.
AC   P34616;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYPOTHETICAL 375.7 KD PROTEIN ZK112.7 IN CHROMOSOME III PRECURSOR.
GN   ZK112.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. STRONG, TO
CC       D.MELANOGASTER FAT TUMOR SUPPRESSOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14324; AAA28182.1; -.
DR   PIR; S44887; S44887.
DR   WORMPEP; ZK112.7; CE00378.
DR   PFAM; PF00028; cadherin; 11.
DR   PROSITE; PS00232; CADHERIN; 8.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW   Hypothetical protein; Cell adhesion; Signal; Transmembrane;
KW   Cytoskeleton; Glycoprotein; Calcium-binding; Repeat.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?   3343       HYPOTHETICAL PROTEIN ZK112.7.
FT   DOMAIN        ?   3228       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   3229   3250       POTENTIAL.
FT   DOMAIN     3251   3343       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     22     22       POTENTIAL.
FT   CARBOHYD    149    149       POTENTIAL.
FT   CARBOHYD    250    250       POTENTIAL.
FT   CARBOHYD    288    288       POTENTIAL.
FT   CARBOHYD    369    369       POTENTIAL.
FT   CARBOHYD    467    467       POTENTIAL.
FT   CARBOHYD    612    612       POTENTIAL.
FT   CARBOHYD    752    752       POTENTIAL.
FT   CARBOHYD    806    806       POTENTIAL.
FT   CARBOHYD    941    941       POTENTIAL.
FT   CARBOHYD    966    966       POTENTIAL.
FT   CARBOHYD    970    970       POTENTIAL.
FT   CARBOHYD    985    985       POTENTIAL.
FT   CARBOHYD   1042   1042       POTENTIAL.
FT   CARBOHYD   1335   1335       POTENTIAL.
FT   CARBOHYD   1425   1425       POTENTIAL.
FT   CARBOHYD   1429   1429       POTENTIAL.
FT   CARBOHYD   1557   1557       POTENTIAL.
FT   CARBOHYD   1563   1563       POTENTIAL.
FT   CARBOHYD   1597   1597       POTENTIAL.
FT   CARBOHYD   1624   1624       POTENTIAL.
FT   CARBOHYD   1695   1695       POTENTIAL.
FT   CARBOHYD   1702   1702       POTENTIAL.
FT   CARBOHYD   1895   1895       POTENTIAL.
FT   CARBOHYD   1900   1900       POTENTIAL.
FT   CARBOHYD   2053   2053       POTENTIAL.
FT   CARBOHYD   2129   2129       POTENTIAL.
FT   CARBOHYD   2203   2203       POTENTIAL.
FT   CARBOHYD   2382   2382       POTENTIAL.
FT   CARBOHYD   2391   2391       POTENTIAL.
FT   CARBOHYD   2410   2410       POTENTIAL.
FT   CARBOHYD   2414   2414       POTENTIAL.
FT   CARBOHYD   2431   2431       POTENTIAL.
FT   CARBOHYD   2527   2527       POTENTIAL.
FT   CARBOHYD   2530   2530       POTENTIAL.
FT   CARBOHYD   2564   2564       POTENTIAL.
FT   CARBOHYD   2621   2621       POTENTIAL.
FT   CARBOHYD   2665   2665       POTENTIAL.
FT   CARBOHYD   2712   2712       POTENTIAL.
FT   CARBOHYD   2798   2798       POTENTIAL.
FT   CARBOHYD   2809   2809       POTENTIAL.
FT   CARBOHYD   2927   2927       POTENTIAL.
FT   CARBOHYD   2976   2976       POTENTIAL.
FT   CARBOHYD   3045   3045       POTENTIAL.
FT   CARBOHYD   3222   3222       POTENTIAL.
FT   CARBOHYD   3225   3225       POTENTIAL.
SQ   SEQUENCE   3343 AA;  375745 MW;  CBC147E7 CRC32;


  Query Match           53.4%;  Score 47;  DB 1;  Length 3343;
  Best Local Similarity 53.3%;  Pred. No. 13;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |  || : |:||| |
Db    213 CHLNILILDINDNPP 227


RESULT  12
CAD8_HUMAN
ID   CAD8_HUMAN     STANDARD;      PRT;   793 AA.
AC   P55286;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CADHERIN-8 PRECURSOR.
GN   CDH8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 95073006.
RA   TANIHARA H., SANO K., HEIMARK R.L., ST JOHN T., SUZUKI S.;
RT   "Cloning of five human cadherins clarifies characteristic features of
RT   cadherin extracellular domain and provides further evidence for two
RT   structurally different types of cadherin.";
RL   Cell Adhes. Commun. 2:15-26(1994).
RN   [2]
RP   SEQUENCE OF 288-793 FROM N.A.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE; 91283540.
RA   SUZUKI S., SANO K., TANIHARA H.;
RT   "Diversity of the cadherin family: evidence for eight new cadherins
RT   in nervous tissue.";
RL   Cell Regul. 2:261-270(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
```

```
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAINLY EXPRESSED IN BRAIN. FOUND IN CERTAIN
CC       NERVE CELL LINES, SUCH AS RETINOBLASTS, GLIOMA CELLS AND
CC       NEUROBLASTS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34060; AAA35628.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 603008; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1      ?       POTENTIAL.
FT   PROPEP       ?     54       POTENTIAL.
FT   CHAIN       55    793       CADHERIN-8.
FT   DOMAIN      55    614       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   615    635       POTENTIAL.
FT   DOMAIN     636    793       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      55    160       CADHERIN 1.
FT   REPEAT     161    269       CADHERIN 2.
FT   REPEAT     270    384       CADHERIN 3.
FT   REPEAT     385    487       CADHERIN 4.
FT   REPEAT     488    609       CADHERIN 5.
FT   CARBOHYD    26     26       POTENTIAL.
FT   CARBOHYD    50     50       POTENTIAL.
FT   CARBOHYD   181    181       POTENTIAL.
FT   CARBOHYD   456    456       POTENTIAL.
FT   CARBOHYD   466    466       POTENTIAL.
FT   CARBOHYD   537    537       POTENTIAL.
SQ   SEQUENCE   793 AA;  87570 MW;  5477875A CRC32;
```

```
  Query Match             52.3%;  Score 46;  DB 1;  Length 793;
  Best Local Similarity   72.7%;  Pred. No. 4.5;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     148 IKVQDINDNAP 158
```

```
RESULT  13
CAD8_MOUSE
ID   CAD8_MOUSE     STANDARD;      PRT;   799 AA.
AC   P97291;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CADHERIN-8 PRECURSOR.
GN   CDH8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER / NIH;
RX   MEDLINE; 97174321.
RA   KOREMATSU K., REDIES C.;
RT   "Restricted expression of cadherin-8 in segmental and functional
RT   subdivisions of the embryonic mouse brain.";
RL   Dev. Dyn. 208:178-189(1997).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95600; CAA64857.1; -.
DR   HSSP; P15116; 1NCI.
DR   MGD; MGI:107434; CDH8.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1      29       POTENTIAL.
FT   PROPEP      30      61       POTENTIAL.
FT   CHAIN       62     799       CADHERIN-8.
FT   DOMAIN      62     621       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   622     642       POTENTIAL.
FT   DOMAIN     643     799       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      62     167       CADHERIN 1.
FT   REPEAT     168     276       CADHERIN 2.
FT   REPEAT     277     391       CADHERIN 3.
FT   REPEAT     392     494       CADHERIN 4.
FT   REPEAT     495     616       CADHERIN 5.
FT   CARBOHYD   188     188       POTENTIAL.
FT   CARBOHYD   463     463       POTENTIAL.
FT   CARBOHYD   473     473       POTENTIAL.
FT   CARBOHYD   544     544       POTENTIAL.
SQ   SEQUENCE   799 AA;  88200 MW;  F1B2625D CRC32;
```

```
  Query Match             52.3%;  Score 46;  DB 1;  Length 799;
  Best Local Similarity   72.7%;  Pred. No. 4.5;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     155 IKVQDINDNAP 165
```

```
RESULT  14
CADE_HUMAN
ID   CADE_HUMAN     STANDARD;      PRT;   790 AA.
AC   Q13634;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CADHERIN-14 PRECURSOR.
GN   CDH14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97184182.
RA   SHIBATA T., SHIMOYAMA Y., GOTOH M., HIROHASHI S.;
RT   "Identification of human cadherin-14, a novel neurally specific type
RT   II cadherin, by protein interaction cloning.";
RL   J. Biol. Chem. 272:5236-5240(1997).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
```

```
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U59325; AAB02933.1; -.
DR   HSSP; P15116; 1NCI.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   PROPEP       25     53       POTENTIAL.
FT   CHAIN        54    790       CADHERIN-14.
FT   DOMAIN       54    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    636       POTENTIAL.
FT   DOMAIN      637    790       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       54    159       CADHERIN 1.
FT   REPEAT      160    268       CADHERIN 2.
FT   REPEAT      269    383       CADHERIN 3.
FT   REPEAT      384    486       CADHERIN 4.
FT   REPEAT      487    608       CADHERIN 5.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    455    455       POTENTIAL.
FT   CARBOHYD    536    536       POTENTIAL.
SQ   SEQUENCE   790 AA;  88072 MW;  4596258F CRC32;


  Query Match          52.3%;  Score 46;  DB 1;  Length 790;
  Best Local Similarity   72.7%;  Pred. No. 4.5;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |||:|:||| |
Db     147 IKVQDINDNAP 157


RESULT  15
DSG2_HUMAN
ID   DSG2_HUMAN     STANDARD;      PRT;  1117 AA.
AC   Q14126;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 2 PRECURSOR (HDGC).
GN   DSG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CARCINOMA;
RX   MEDLINE; 94192736.
RA   SCHAEFER S., KOCH P.J., FRANKE W.W.;
RT   "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT   expression catalogue of the desmoglein subfamily of desmosomal
RT   cadherins.";
RL   Exp. Cell Res. 211:391-399(1994).
RN   [2]
RP   SEQUENCE OF 777-1117 FROM N.A.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z26317; CAA81226.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 125671; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     48       POTENTIAL.
FT   CHAIN        49   1117       DESMOGLEIN 2.
FT   DOMAIN       49    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    633       POTENTIAL.
FT   DOMAIN      634   1117       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       49    159       CADHERIN 1.
FT   REPEAT      160    272       CADHERIN 2.
FT   REPEAT      273    387       CADHERIN 3.
FT   REPEAT      388    502       CADHERIN 4.
FT   REPEAT      880    911       DESMOGLEIN REPEAT 1.
FT   REPEAT      912    941       DESMOGLEIN REPEAT 2.
FT   REPEAT      942    967       DESMOGLEIN REPEAT 3.
FT   REPEAT      968    991       DESMOGLEIN REPEAT 4.
FT   REPEAT      992   1020       DESMOGLEIN REPEAT 5.
FT   REPEAT     1021   1050       DESMOGLEIN REPEAT 6.
FT   CARBOHYD    111    111       POTENTIAL.
FT   CARBOHYD    181    181       POTENTIAL.
FT   CARBOHYD    308    308       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    513    513       POTENTIAL.
SQ   SEQUENCE   1117 AA;  122385 MW;  84D3B898 CRC32;


  Query Match          52.3%;  Score 46;  DB 1;  Length 1117;
  Best Local Similarity   64.3%;  Pred. No. 6.4;
  Matches    9;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 ECNIKVKDVNDNFP 15
           |  ||| |:||| |
Db     143 ELRIKVLDINDNEP 156


Search completed: January  7, 2000, 13:25:40
Job time: 392 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:22 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-5
Perfect score: 88
Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	82	93.2	993	11	O35902	O35902 mus musculu
2	57	64.8	519	13	Q90425	Q90425 brachydanio
3	52	59.1	832	4	Q15336	Q15336 homo sapien
4	52	59.1	832	4	Q12864	Q12864 homo sapien
5	51	58.0	2163	5	O01912	O01912 caenorhabdi
6	50	56.8	1507	5	Q24298	Q24298 drosophila
7	50	56.8	1180	11	O55134	O55134 mus musculu
8	49	55.7	222	3	O13737	O13737 schizosacch
9	49	55.7	810	4	O75279	O75279 homo sapien
10	49	55.7	807	4	O75288	O75288 homo sapien
11	49	55.7	936	4	O95206	O95206 homo sapien
12	49	55.7	915	5	O62328	O62328 caenorhabdi
13	48	54.5	824	4	O75287	O75287 homo sapien
14	48	54.5	784	11	O35542	O35542 mus musculu
15	48	54.5	947	11	O88689	O88689 mus musculu
16	47	53.4	814	6	O77704	O77704 canis famil
17	47	53.4	898	11	O88193	O88193 mus musculu
18	46	52.3	889	4	O14917	O14917 homo sapien
19	46	52.3	4590	4	Q14517	Q14517 homo sapien
20	46	52.3	816	4	O75284	O75284 homo sapien
21	46	52.3	799	11	O54800	O54800 rattus norv
22	46	52.3	532	11	O54801	O54801 rattus norv
23	45	51.1	934	4	O60622	O60622 homo sapien
24	45	51.1	134	4	O15202	O15202 homo sapien
25	45	51.1	904	4	Q08192	Q08192 homo sapien
26	45	51.1	814	4	O75281	O75281 homo sapien
27	45	51.1	824	4	O75286	O75286 homo sapien
28	45	51.1	829	4	O75309	O75309 homo sapien
29	45	51.1	3380	5	Q24292	Q24292 drosophila
30	45	51.1	578	11	Q63423	Q63423 rattus norv
31	45	51.1	895	11	O55122	O55122 mus musculu
32	45	51.1	932	13	O13129	O13129 gallus gall
33	45	51.1	148	13	O93300	O93300 brachydanio
34	45	51.1	979	13	O93509	O93509 xenopus lae
35	44	50.0	235	2	O67056	O67056 aquifex aeo
36	44	50.0	1026	4	Q08174	Q08174 homo sapien
37	44	50.0	842	4	O15053	O15053 homo sapien
38	44	50.0	807	4	O75277	O75277 homo sapien
39	44	50.0	792	4	O75278	O75278 homo sapien
40	44	50.0	844	4	O75280	O75280 homo sapien
41	44	50.0	789	4	O75282	O75282 homo sapien
42	44	50.0	620	5	O01909	O01909 caenorhabdi
43	44	50.0	2809	5	O61230	O61230 lytechinus
44	44	50.0	3097	5	O15943	O15943 drosophila
45	44	50.0	829	6	Q28634	Q28634 oryctolagus

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 93.2%; Score 82; DB 11; Length 993;
Best Local Similarity 86.7%; Pred. No. 3.2e-05;
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CECNIKVKDVNDNFP 15
           |||:||:|||||||
Db     251 CECSIKIKDVNDNFP 265
```

RESULT 2
Q90425
ID Q90425 PRELIMINARY; PRT; 519 AA.
AC Q90425;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE VENTRAL NEURAL CADHERIN (FRAGMENT).
GN VNC.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinoidea; Cyprinidae; Rasborinae; Danio.

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FRANKLIN J.L., SARGENT T.D.;
RL   Dev. Dyn. 206:0-0(0).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; U41419; AAB47406.1; -.
DR   PFAM; PF00028; cadherin; 3.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 1.
DR   ZFIN; ZDB-GENE-980526-170; vnc.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT   NON_TER       1      1
SQ   SEQUENCE   519 AA;  57807 MW;  E6DA0079 CRC32;


  Query Match             64.8%;  Score 57;  DB 13;  Length 519;
  Best Local Similarity   91.7%;  Pred. No. 0.16;
  Matches   11;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 NIKVKDVNDNFP 15
          |||||||||| |
Db    202 NIKVKDVNDNAP 213


RESULT   3
Q15336
ID   Q15336      PRELIMINARY;      PRT;   832 AA.
AC   Q15336;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   LI-CADHERIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BOETTINGER A., KREFT B., FIEGER C., DLOUHY B., BERNDORFF D.,
RA   GOESSNER R., TAUBER R.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; X83228; CAA58231.1; -.
DR   PFAM; PF00028; cadherin; 7.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   832 AA;  92207 MW;  65C92D4C CRC32;


  Query Match             59.1%;  Score 52;  DB 4;  Length 832;
  Best Local Similarity   64.3%;  Pred. No. 1.6;
  Matches    9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ECNIKVKDVNDNFP 15
          | ::|||||||| |
Db    325 EIHVKVKDINDNPP 338


RESULT   4
Q12864
ID   Q12864      PRELIMINARY;      PRT;   832 AA.
AC   Q12864;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   INTESTINAL PEPTIDE-ASSOCIATED TRANSPORTER HPT-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RX   MEDLINE; 94204643.
RA   DANTZIG A.H., HOSKINS J., TABAS L.B., BRIGHT S., SHEPARD R.L.,
RA   JENKINS I.L., DUCKWORTH D.C., SPORTSMAN R., MACKENSEN D.,
RA   ROSTECK P.R., SKATRUD P.L.;
RT   "Association of intestinal peptide transport with a protein related
RT   to the cadherin superfamily.";
RL   Science 264:430-433(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; U07969; AAA19021.1; -.
DR   PFAM; PF00028; cadherin; 7.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   832 AA;  92147 MW;  B5193483 CRC32;


  Query Match             59.1%;  Score 52;  DB 4;  Length 832;
  Best Local Similarity   64.3%;  Pred. No. 1.6;
  Matches    9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 ECNIKVKDVNDNFP 15
          | ::|||||||| |
Db    325 EIHVKVKDINDNPP 338


RESULT   5
O01912
ID   O01912      PRELIMINARY;      PRT;  2163 AA.
AC   O01912;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS.
GN   R10F2.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   DU Z., GATTUNG S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF003388; AAB54266.1; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PFAM; PF00028; cadherin; 15.
DR   PROSITE; PS00232; CADHERIN; 8.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   2163 AA;  238609 MW;  6C85C652 CRC32;


  Query Match             58.0%;  Score 51;  DB 5;  Length 2163;
  Best Local Similarity   53.3%;  Pred. No. 5.9;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       1 CECNIKVKDVNDNFP 15
           |:|:: | | ||| |
Db    1629 CKCHVIVLDENDNLP 1643


RESULT   6
Q24298
ID   Q24298      PRELIMINARY;      PRT;  1507 AA.
AC   Q24298;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   DE-CADHERIN PRECURSOR (SHOTGUN) (SHG).
GN   SHG OR GP150 OR DE-CADHERIN.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 95046887.
RA   ODA H., UEMURA T., HARADA Y., IWAI Y., TAKEICHI M.;
RT   "A Drosophila homolog of cadherin associated with armadillo and
RT   essential for embryonic cell-cell adhesion.";
RL   Dev. Biol. 165:716-726(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC       N-CADHERIN MAY BE INVOLVED IN NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: STAGE 10 EMBRYOS EXHIBIT INTENSE EXPRESSION
CC       IN EPITHELIAL CELLS.
CC       STAGE 14 EMBRYOS SHOW EXPRESSION IN THE HINDGUT
CC       (AT THE APICAL POLES OF CELL-CELL BOUNDARIES), AT THE APICAL
CC       JUNCTIONS OF TRACHEAL CELLS AND IN THE DORSAL LONGITUDINAL TRUNK.
CC       IN STAGE 16 EMBRYOS THE GLIAL MIDLINE CELLS OF THE CENTRAL
CC       NERVOUS SYSTEM SHOW STRONG EXPRESSION.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
DR   EMBL; D28749; BAA05942.1; -.
DR   PROSITE; PS00232; CADHERIN; 5.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   FLYBASE; FBgn0003391; shg.
DR   PFAM; PF00028; cadherin; 7.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PFAM; PF00054; laminin_G; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; Zymogen.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?    261       POTENTIAL.
FT   CHAIN       262   1507       DE-CADHERIN PROTEIN.
FT   DOMAIN      262   1328       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1329   1349       POTENTIAL.
FT   DOMAIN     1350   1507       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1112   1123       EGF-LIKE.
FT   REPEAT      302    412       CADHERIN 1 (CR1).
FT   REPEAT      413    522       CADHERIN 2 (CR2).
FT   REPEAT      523    623       CADHERIN 3 (CR3).
FT   REPEAT      624    733       CADHERIN 4 (CR4).
FT   REPEAT      734    836       CADHERIN 5 (CR5).
FT   CARBOHYD    317    317       POTENTIAL.
FT   CARBOHYD    466    466       POTENTIAL.
FT   CARBOHYD    552    552       POTENTIAL.
FT   CARBOHYD    766    766       POTENTIAL.
FT   CARBOHYD    949    949       POTENTIAL.
FT   CARBOHYD    983    983       POTENTIAL.
FT   CARBOHYD    999    999       POTENTIAL.
FT   CARBOHYD   1073   1073       POTENTIAL.
FT   CARBOHYD   1145   1145       POTENTIAL.
FT   CARBOHYD   1274   1274       POTENTIAL.
FT   CARBOHYD   1290   1290       POTENTIAL.
SQ   SEQUENCE   1507 AA;  169854 MW;  911CC772 CRC32;


  Query Match             56.8%;  Score 50;  DB 5;  Length 1507;
  Best Local Similarity   46.7%;  Pred. No. 6;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |  |: ::|:||| |
Db     179 CTFNVTIEDINDNAP 193


RESULT   7
O55134
ID   O55134      PRELIMINARY;      PRT;  1180 AA.
AC   O55134;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROTOCADHERIN-4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN CAPILLARY;
RA   TELO P.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; Y08715; CAA69965.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1180 AA;  128672 MW;  48C0883F CRC32;


  Query Match             56.8%;  Score 50;  DB 11;  Length 1180;
  Best Local Similarity   66.7%;  Pred. No. 4.7;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |:  ||| ||||| |
Db     336 CKVLIKVLDVNDNAP 350


RESULT   8
O13737
ID   O13737      PRELIMINARY;      PRT;   222 AA.
AC   O13737;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 24.9 KD PROTEIN C16E8.02 IN CHROMOSOME I.
GN   SPAC16E8.02.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST YGL010W AND SOME, TO N.CRASSA ATP-6.
DR   EMBL; Z98529; CAB11031.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     65     85       POTENTIAL.
FT   TRANSMEM    105    125       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
```

```
FT   DOMAIN       30      33       POLY-LEU.
SQ   SEQUENCE   222 AA;  24993 MW;  396C437E CRC32;


  Query Match             55.7%;  Score 49;  DB 3;  Length 222;
  Best Local Similarity   66.7%;  Pred. No. 1.4;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 NIKVKDVNDNFP 15
          |||:||| :|:|
Db    167 NIKLKDVGENYP 178


RESULT   9
O75279
ID   O75279      PRELIMINARY;      PRT;   810 AA.
AC   O75279;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KIAA0345-LIKE 3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KIMMERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
RA   KADNER K., MIGUEL T., MILLER C., PITLUCK S., POLLARD M., ROJESKI H.,
RA   SUBRAMANIAN S., MARTIN C.H.;
RT   "Sequencing of human chromosome 5.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   RICKE D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AC005609; AAC34316.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   810 AA;  88391 MW;  9F4E3C7D CRC32;


  Query Match             55.7%;  Score 49;  DB 4;  Length 810;
  Best Local Similarity   66.7%;  Pred. No. 4.7;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 NIKVKDVNDNFP 15
          |::|||:||| |
Db    120 NVEVKDINDNPP 131


RESULT  10
O75288
ID   O75288      PRELIMINARY;      PRT;   807 AA.
AC   O75288;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KIAA0345-LIKE 13.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KIMMERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
RA   KADNER K., MIGUEL T., MILLER C., PITLUCK S., POLLARD M., ROJESKI H.,
RA   SUBRAMANIAN S., MARTIN C.H.;
RT   "Sequencing of human chromosome 5.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   RICKE D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AC005609; AAC34325.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   807 AA;  87639 MW;  B0523965 CRC32;


  Query Match             55.7%;  Score 49;  DB 4;  Length 807;
  Best Local Similarity   60.0%;  Pred. No. 4.7;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |:  :|| ||||| |
Db    334 CKVLVKVLDVNDNAP 348


RESULT  11
O95206
ID   O95206      PRELIMINARY;      PRT;   936 AA.
AC   O95206;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROTOCADHERIN.
GN   PCDH8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99005535.
RA   STREHL S., GLATT K., LIU Q.M., GLATT H., LALANDE M.;
RT   "Characterization of two novel protocadherins (PCDH8 and PCDH9)
RT   localized on human chromosome 13 and mouse chromosome 14.";
RL   Genomics 53:81-89(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF061573; AAC70009.1; -.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   936 AA;  98879 MW;  3A6F1EC8 CRC32;


  Query Match             55.7%;  Score 49;  DB 4;  Length 936;
  Best Local Similarity   46.7%;  Pred. No. 5.4;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CECNIKVKDVNDNFP 15
          |:  ::::||||| |
Db    338 CKVIVRIRDVNDNAP 352


RESULT  12
O62328
ID   O62328      PRELIMINARY;      PRT;   915 AA.
AC   O62328;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   R05H10.6 PROTEIN.
GN   R05H10.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MCMURRAY A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; Z83119; CAB05582.1; -.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   915 AA;  101482 MW;  A25FF734 CRC32;


  Query Match             55.7%;  Score 49;  DB 5;  Length 915;
  Best Local Similarity   81.8%;  Pred. No. 5.3;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           :|||||||| |
Db     625 VKVKDVNDNGP 635


RESULT  13
O75287
ID   O75287      PRELIMINARY;      PRT;   824 AA.
AC   O75287;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KIAA0345-LIKE 12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KIMMERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
RA   KADNER K., MIGUEL T., MILLER C., PITLUCK S., POLLARD M., ROJESKI H.,
RA   SUBRAMANIAN S., MARTIN C.H.;
RT   "Sequencing of human chromosome 5.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   RICKE D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AC005609; AAC34324.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   824 AA;  88866 MW;  56C50E1B CRC32;


  Query Match             54.5%;  Score 48;  DB 4;  Length 824;
  Best Local Similarity   46.7%;  Pred. No. 6.9;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |: ::|: |:||| |
Db     334 CKISLKLVDINDNTP 348


RESULT  14
O35542
ID   O35542      PRELIMINARY;      PRT;   784 AA.
AC   O35542;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CADHERIN-5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97364256.
RA   MATSUYOSHI N., TODA K., HORIGUCHI Y., TANAKA T., NAKAGAWA S.,
RA   TAKEICHI M., IMAMURA S.;
RT   "In vivo evidence of the critical role of cadherin-5 in murine
RT   vascular integrity.";
RL   Proc. Assoc. Am. Physicians 109:362-371(1997).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D63942; BAA22617.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   784 AA;  87902 MW;  68D4FB78 CRC32;


  Query Match             54.5%;  Score 48;  DB 11;  Length 784;
  Best Local Similarity   63.6%;  Pred. No. 6.6;
  Matches    7;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 IKVKDVNDNFP 15
           |:::|:|||||
Db     244 IRLEDINDNFP 254


RESULT  15
O88689
ID   O88689      PRELIMINARY;      PRT;   947 AA.
AC   O88689;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CNR1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98318234.
RA   KOHMURA N., SENZAKI K., HAMADA S., KAI N., YASUDA R., WATANABE M.,
RA   ISHII H., YASUDA M., MISHINA M., YAGI T.;
RT   "Diversity revealed by a novel family of cadherins expressed in
RT   neurons at a synaptic complex.";
RL   Neuron 20:1137-1151(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D86916; BAA29045.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   947 AA;  103143 MW;  AB588B2C CRC32;


  Query Match             54.5%;  Score 48;  DB 11;  Length 947;
  Best Local Similarity   46.7%;  Pred. No. 7.9;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CECNIKVKDVNDNFP 15
           |:  ::|:|:||| |
Db     334 CKVIVQVEDINDNVP 348
```

Search completed: January 12, 2000, 23:15:23
Job time: 186 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:14:50 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-1
Perfect score: 74
Sequence: 1 ATQKITYRISGVGID 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	999	1	IJHUG3	desmoglein 3 precu
2	64	86.5	1049	1	IJHUG1	desmoglein 1 precu
3	64	86.5	1043	1	IJBOG1	desmoglein 1 precu
4	44	59.5	388	2	T10963	phosphoribosylform
5	42	56.8	552	1	A55760	cystathionine beta
6	42	56.8	897	2	A39405	beta-galactosidase
7	40	54.1	2023	2	T13154	polycomb protein e
8	39	52.7	441	2	T01000	hypothetical prote
9	39	52.7	504	2	JC5830	myocilin - human
10	38	51.4	956	1	A46016	thrombospondin 3 -
11	38	51.4	1177	1	D71416	probable PDR5-like
12	38	51.4	325	2	I57009	3-Hydroxy-3-methyl
13	38	51.4	298	2	B45470	hydroxymethylgluta
14	38	51.4	433	2	T02655	hydroxymethylgluta
15	38	51.4	331	2	E70367	phosphoribosylform
16	38	51.4	147	2	A69198	hypothetical prote
17	38	51.4	177	2	S65159	hypothetical prote
18	37	50.0	325	1	A45470	hydroxymethylgluta
19	37	50.0	561	1	A42790	cystathionine beta
20	37	50.0	478	1	C42790	cystathionine beta
21	37	50.0	1161	1	S31213	nidogen precursor
22	37	50.0	1363	1	VGIHNM	E2 glycoprotein pr
23	37	50.0	1363	1	VGIHQU	E2 glycoprotein pr
24	37	50.0	1363	1	VGIHF1	E2 glycoprotein pr
25	37	50.0	1363	1	VGIHL9	E2 glycoprotein pr
26	37	50.0	1363	1	VGIHVA	E2 glycoprotein pr
27	37	50.0	1363	1	VGIHLY	E2 glycoprotein pr
28	37	50.0	546	2	B42790	cystathionine beta
29	37	50.0	356	2	D42790	cystathionine beta
30	37	50.0	547	2	JX0145	hemoprotein H-450
31	37	50.0	1117	2	S38673	desmoglein 2 - hum
32	37	50.0	1363	2	S44240	surface protein -
33	37	50.0	1363	2	S44241	surface protein -
34	37	50.0	1218	2	S38182	probable transport
35	37	50.0	1121	2	S57058	probable membrane
36	37	50.0	402	2	E71279	hypothetical prote
37	36	48.6	740	1	JC6010	formate C-acetyltr
38	36	48.6	355	1	JQ2256	phosphoribosylform
39	36	48.6	712	1	IJMSCT	T-cadherin precurs
40	36	48.6	323	1	H64130	glycosyl transfera
41	36	48.6	690	2	A47298	NADPH--ferrihemopr
42	36	48.6	762	2	C69657	cobalamin-independ
43	36	48.6	125	2	T06719	phosphoribosylform
44	36	48.6	82	2	I51136	MHC class II beta
45	36	48.6	82	2	I51135	MHC class II beta

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 74; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGID 15
||||||||||||||||
Db 79 ATQKITYRISGVGID 93

RESULT 2
IJHUG1
desmoglein 1 precursor - human
N;Alternate names: desmosomal glycoprotein I
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S16906; A39706; A61254; A61279; S16158
R;Buxton, R.S.
submitted to the EMBL Data Library, November 1990
A;Reference number: S16906
A;Accession: S16906
A;Molecule type: mRNA

A;Residues: 1-1049 <BUX>
A;Cross-references: EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
R;Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; R
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991
A;Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,
A;Reference number: A39706; MUID:91271279
A;Accession: A39706
A;Molecule type: mRNA
A;Residues: 24-1049 <WHE>
A;Cross-references: GB:X56654
R;Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
J. Cell Sci. 99, 809-821, 1991
A;Title: Structural analysis and expression of human desmoglein: a cadherin-like compone
A;Reference number: A61254; MUID:92121251
A;Accession: A61254
A;Molecule type: mRNA
A;Residues: 26-1049 <NIL>
R;Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee
Biochem. Soc. Trans. 19, 1060-1064, 1991
A;Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfami
A;Reference number: A61279; MUID:92175187
A;Accession: A61279
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-55 <WH3>
C;Genetics:
A;Gene: GDB:DSG1
A;Cross-references: GDB:126563; OMIM:125670
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1049/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-493/Domain: cadherin repeat homology <CR4>
F;509-530/Region: serine/threonine-rich
F;549-569/Domain: transmembrane #status predicted <TMM>
F;572-1049/Domain: intracellular #status predicted <INT>
F;840-869/Domain: desmoglein repeat <DG1>
F;870-899/Domain: desmoglein repeat <DG2>
F;900-927/Domain: desmoglein repeat <DG3>
F;928-956/Domain: desmoglein repeat <DG4>
F;969-1019/Region: glycine/serine-rich
F;110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.5%; Score 64; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 0.0014;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGID 15
| |::|||||||||||
Db 79 ANQQVTYRISGVGID 93

RESULT 3
IJBOG1
desmoglein 1 precursor - bovine
N;Alternate names: desmoglein BDGM
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S14603; A38872; A37785; S38721; A48173; S24412
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
submitted to the EMBL Data Library, March 1991
A;Description: Complete sequence of the desmoglein precursor and evidence for the existe
A;Reference number: S14603
A;Accession: S14603
A;Molecule type: mRNA
A;Residues: 1-1043 <KOC>
A;Cross-references: EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti
A;Reference number: A38872; MUID:92037656
A;Accession: A38872
A;Molecule type: mRNA
A;Residues: 1-87;968-1043 <KO2>
A;Cross-references: GB:S64268; GB:S64270
R;Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990
A;Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion
A;Reference number: A37785; MUID:91097553
A;Accession: A37785
A;Molecule type: mRNA
A;Residues: 44-123,'V',125-493 <GOO>
A;Cross-references: GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
R;Zimbelmann, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S38721
A;Accession: S38721
A;Molecule type: mRNA
A;Residues: 44-1043 <ZIM>
A;Cross-references: EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
R;Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.
Eur. J. Cell Biol. 53, 1-12, 1990
A;Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m
A;Reference number: A48173; MUID:91168965
A;Accession: A48173
A;Molecule type: mRNA
A;Residues: 44-1001,'AQPPSAT' <KO3>
A;Cross-references: GB:X57784
A;Note: this sequence has been revised in references A38872 and S38721
C;Genetics:
A;Gene: DSG1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1043/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-491/Domain: cadherin repeat homology <CR4>
F;549-574/Domain: transmembrane #status predicted <TMM>
F;575-1043/Domain: intracellular #status predicted <INT>
F;846-875/Domain: desmoglein repeat <DG1>
F;876-905/Domain: desmoglein repeat <DG2>
F;906-933/Domain: desmoglein repeat <DG3>
F;934-962/Domain: desmoglein repeat <DG4>
F;963-1012/Region: glycine/serine-rich
F;110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.5%; Score 64; DB 1; Length 1043;
Best Local Similarity 80.0%; Pred. No. 0.0014;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGID 15
| |::|||||||||||
Db 79 ANQQVTYRISGVGID 93

RESULT 4
T10963
phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1) - cowpea
N;Alternate names: AIR synthase; AIRS; phosphoribosylaminoimidazole synthetase
C;Species: Vigna unguiculata (cowpea)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Sep-1999
C;Accession: T10963
R;Smith, P.M.; Mann, A.J.; Goggin, D.E.; Atkins, C.A.

Plant Mol. Biol. 36, 811-820, 1998
A;Title: AIR synthetase in cowpea nodules: a single gene product targeted to two organel
A;Reference number: Z17233; MUID:98179096
A;Accession: T10963
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-388 <SMI>
A;Cross-references: EMBL:U30895; NID:g945059; PID:g945060
A;Experimental source: strain Vita 3; root nodule; clone pAIRS19
C;Genetics:
A;Gene: pur5
C;Function:
A;Description: involved in the fifth step of the purine biosynthetic pathway, converting
C;Superfamily: phosphoribosylformylglycinamidine cyclo-ligase; phosphoribosylformylglyci
C;Keywords: cyclo-ligase; purine nucleotide biosynthesis

Query Match 59.5%; Score 44; DB 2; Length 388;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGID 15
|:| :||| :|| ||
Db 62 ASQGLTYRDAGVDID 76

RESULT 5
A55760
cystathionine beta-synthase (EC 4.2.1.22) - human
N;Alternate names: beta-thionase; methylcysteine synthase; serine sulfhydrase
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A55760; S49459; I54342
R;Kruger, W.D.; Cox, D.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 6614-6618, 1994
A;Title: A yeast system for expression of human cystathionine beta-synthase: structural
A;Reference number: A55760; MUID:94294429
A;Accession: A55760
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-552 <KRU>
A;Cross-references: GB:L14577
R;Chasse, J.F.; Paris, D.; Paly, E.; Kamoun, P.; London, J.
submitted to the EMBL Data Library, October 1994
A;Description: Characterization of human cystathionine beta-synthase (CBS) cDNAs: eviden
A;Reference number: S49459
A;Accession: S49459
A;Molecule type: mRNA
A;Residues: 1-360,'A',362-364,366-418,'G',420-492,'L',494-498,'R',500-543,'V',545-552 <C
A;Cross-references: EMBL:X82166; NID:g558581; PIDN:CAA57656.1; PID:g558582
R;Kraus, J.P.; Le, K.; Swaroop, M.; Ohura, T.; Tahara, T.; Rosenberg, L.E.; Roper, M.D.;
Hum. Mol. Genet. 2, 1633-1638, 1993
A;Title: Human cystathionine beta-synthase cDNA: sequence, alternative splicing and expr
A;Reference number: I54342; MUID:94093551
A;Accession: I54342
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-360,'A',362-364,366-418,'G',420-492,'L',494-498,'R',500-543,'V',545-552 <R
A;Cross-references: GB:L19501; NID:g388715; PIDN:AAA19874.1; PID:g388716
C;Genetics:
A;Gene: GDB:CBS
A;Cross-references: GDB:119754; OMIM:236200
A;Map position: 21q22.3-21q22.3
C;Superfamily: cystathionine beta-synthase; CBS homology
C;Keywords: carbon-oxygen lyase; hydro-lyase
F;422-470/Domain: CBS homology <CBS>

Query Match 56.8%; Score 42; DB 1; Length 552;
Best Local Similarity 42.9%; Pred. No. 7.4;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TQKITYRISGVGID 15
|:: || : |:| |
Db 296 TEQTTYEVEGIGYD 309

RESULT 6
A39405
beta-galactosidase (EC 3.2.1.23) - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 23-Jun-1993
C;Accession: A39405
R;Hancock, K.R.; Rockman, E.; Young, C.A.; Pearce, L.; Maddox, I.S.; Scott, D.B.
J. Bacteriol. 173, 3084-3095, 1991
A;Title: Expression and nucleotide sequence of the Clostridium acetobutylicum beta-ga
A;Reference number: A39405; MUID:91216979
A;Accession: A39405
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-897 <HAN>
A;Cross-references: GB:M35107
C;Keywords: glycosidase; hydrolase

Query Match 56.8%; Score 42; DB 2; Length 897;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 KITYRISGVGI 14
|||| :|| ||
Db 828 KITYEVSGEGI 838

RESULT 7
T13154
polycomb protein enhancer - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: T13154
R;Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.
Development 125, 4055-4066, 1998
A;Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein cons
A;Reference number: Z17611
A;Accession: T13154
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2023 <STA>
A;Cross-references: EMBL:AF079764; NID:g3757889; PID:g3757890; PIDN:AAC64271.1
A;Experimental source: imaginal disc
C;Genetics:
A;Gene: E(Pc)
A;Map position: 2

Query Match 54.1%; Score 40; DB 2; Length 2023;
Best Local Similarity 50.0%; Pred. No. 70;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGI 14
|:||:|| | |:
Db 1453 ASQKVTYAIQKAGV 1466

RESULT 8
T01000
hypothetical protein T5I7.1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 14-May-1999
C;Accession: T01000
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, November 1997
A;Description: Arabidopsis thaliana chromosome II BAC T5I7 genomic sequence.
A;Reference number: Z14159
A;Accession: T01000
A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-441 <ROU>
A;Cross-references: EMBL:AC003000; NID:g2642152; PID:g2642153
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
A;Note: T5I7.1

Query Match 52.7%; Score 39; DB 2; Length 441;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 QKITYRISGVG 13
||: ||::| |
Db 368 QKLLYRVNGAG 378

RESULT 9
JC5830
myocilin - human
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 17-Mar-1999
C;Accession: JC5830
R;Kubota, R.; Kudoh, J.; Mashima, Y.; Asakawa, S.; Minoshima, S.; Hejtmancik, J.F.; Oguc
Biochem. Biophys. Res. Commun. 242, 396-400, 1998
A;Title: Genomic organization of the human myocilin gene (MYOC) responsible for primary
A;Reference number: JC5830; MUID:98113364
A;Accession: JC5830
A;Molecule type: DNA
A;Residues: 1-504 <KUB>
A;Cross-references: DDBJ:AB006686
C;Comment: This cytoskeletal protein is involved in the morphogenesis of the basal body
ucoma.
C;Genetics:
A;Gene: myoc
A;Introns: 202/1; 244/1

Query Match 52.7%; Score 39; DB 2; Length 504;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TQKITYRISGVGID 15
||: |:|| || |
Db 281 TQETTWRIDTVGTD 294

RESULT 10
A46016
thrombospondin 3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A46016; A44124; I55398; S68788; S72433
R;Bornstein, P.; Devarayalu, S.; Edelhoff, S.; Disteche, C.M.
Genomics 15, 607-613, 1993
A;Title: Isolation and characterization of the mouse thrombospondin 3 (Thbs3) gene.
A;Reference number: A46016; MUID:93224149
A;Accession: A46016
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-956 <BOR>
A;Cross-references: GB:L04302; NID:g202200; PIDN:AAA40497.1; PID:g202201
A;Note: sequence extracted from NCBI backbone (NCBIN:129415, NCBIP:129416)
R;Vos, H.L.; Devarayalu, S.; de Vries, Y.; Bornstein, P.
J. Biol. Chem. 267, 12192-12196, 1992
A;Title: Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene family.
A;Reference number: A44124; MUID:92291102
A;Accession: A44124
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 517-956 <VOS>
A;Cross-references: GB:M86620; NID:g201987; PIDN:AAA40430.1; PID:g201989
A;Experimental source: BALB/c
A;Note: sequence extracted from NCBI backbone (NCBIN:106634, NCBIN:106636, NCBIN:1066
)
R;Qabar, A.N.; Lin, Z.; Wolf, F.W.; O'Shea, K.S.; Lawler, J.; Dixit, V.M.
J. Biol. Chem. 269, 1262-1269, 1994
A;Title: Thrombospondin 3 is a developmentally regulated heparin binding protein.
A;Reference number: I55398; MUID:94117438
A;Accession: I55398
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-432,'F',434-562,'D',564-719,'E',721-870,'W',872-956 <RES>
A;Cross-references: GB:L24434; NID:g402718; PIDN:AAA40433.1; PID:g402719
R;Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A;Title: Expression and initial characterization of recombinant mouse thrombospondin
A;Reference number: S68787; MUID:96234006
A;Accession: S68788
A;Molecule type: protein
A;Residues: 'X',23-28,'X',30,'X',32 <CHE>
R;Collins, M.; Bornstein, P.
Nucleic Acids Res. 24, 3661-3669, 1996
A;Title: SP1-binding elements, within the common metaxin-thrombospondin 3 intergenic
A;Reference number: S72433; MUID:97025352
A;Accession: S72433
A;Molecule type: DNA
A;Residues: 1-26 <COL>
A;Cross-references: EMBL:U66257; NID:g1575552; PIDN:AAC52819.1; PID:g1575554
C;Genetics:
A;Gene: Thbs-3
C;Complex: homopentamer, disulfide linked
C;Superfamily: thrombospondin 3; EGF homology
C;Keywords: calcium binding; glycoprotein; homopentamer
F;310,407,644,937/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.4%; Score 38; DB 1; Length 956;
Best Local Similarity 87.5%; Pred. No. 72;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 RISGVGID 15
|:||||||
Db 352 RVSGVGID 359

RESULT 11
D71416
probable PDR5-like ABC transporter - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: D71416
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A;Reference number: A71400; MUID:98121113
A;Accession: D71416
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1177 <BEV>
A;Cross-references: GB:Z97338; NID:g2244870; PID:e326915; PID:g2244881
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: Arabidopsis thaliana probable PDR5-like ABC transporter; ATP-binding c
C;Keywords: ATP
F;439-653/Domain: ATP-binding cassette homology <ABC2>

Query Match 51.4%; Score 38; DB 1; Length 1177;

```
  Best Local Similarity   69.2%;  Pred. No. 90;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 QKITYRISGVGID 15
          |||  ||  ||||
Db   1042 QKIRKRIDEVGID 1054


RESULT  12
I57009
3-Hydroxy-3-methylglutaryl coenzyme A lyase - mouse
C;Species: Mus sp. (mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999
C;Accession: I57009
R;Wang, S.; Nadeau, J.H.; Duncan, A.; Robert, M.F.; Fontaine, G.; Schappert, K.; Johnson
Mamm. Genome 4, 382-387, 1993
A;Title: 3-Hydroxy-3-methylglutaryl coenzyme A lyase (HL): cloning and characterization
A;Reference number: I57009; MUID:93364116
A;Accession: I57009
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-325 <RES>
A;Cross-references: GB:S65036; NID:g409498; PIDN:AAB27965.1; PID:g409499
C;Genetics:
A;Gene: Hmgcl
C;Superfamily: hydroxymethylglutaryl-CoA lyase

  Query Match             51.4%;  Score 38;  DB 2;  Length 325;
  Best Local Similarity   42.9%;  Pred. No. 22;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db    277 ATEDLVYMLNGLGI 290


RESULT  13
B45470
hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 25-Mar-1998
C;Accession: B45470
R;Mitchell, G.A.; Robert, M.F.; Hruz, P.W.; Wang, S.; Fontaine, G.; Behnke, C.E.; Mende-
J. Biol. Chem. 268, 4376-4381, 1993
A;Title: 3-Hydroxy-3-methylglutaryl coenzyme A lyase (HL). Cloning of human and chicken
A;Reference number: A45470; MUID:93179448
A;Accession: B45470
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-298 <MIT>
A;Cross-references: GB:L07034
C;Superfamily: hydroxymethylglutaryl-CoA lyase
C;Keywords: carbon-carbon lyase; mitochondrion; oxo-acid-lyase

  Query Match             51.4%;  Score 38;  DB 2;  Length 298;
  Best Local Similarity   42.9%;  Pred. No. 20;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db    250 ATEDLVYMLNGLGI 263


RESULT  14
T02655
hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4) - Arabidopsis thaliana
N;Alternate names: protein F12C20.16
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 22-Jun-1999
C;Accession: T02655
R;Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sy
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.
A;Reference number: Z14685
A;Accession: T02655
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-433 <ROU>
A;Cross-references: EMBL:AC005168; NID:g3426033; PIDN:AAC32247.1; PID:g3426048
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
A;Introns: 55/2; 124/3; 184/3; 211/3; 266/2; 312/1; 350/3
A;Note: F12C20.16
C;Superfamily: hydroxymethylglutaryl-CoA lyase
C;Keywords: carbon-carbon lyase; oxo-acid-lyase

  Query Match             51.4%;  Score 38;  DB 2;  Length 433;
  Best Local Similarity   35.7%;  Pred. No. 31;
  Matches    5;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:|:
Db    377 ATEDVVYMLNGLGV 390


RESULT  15
E70367
phosphoribosylformylglycinamidine cyclo-ligase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C;Accession: E70367
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: E70367
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-331 <AQF>
A;Cross-references: GB:AE000706; NID:g2983327; PIDN:AAC06918.1; PID:g2983329; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: purM
C;Superfamily: phosphoribosylformylglycinamidine cyclo-ligase; phosphoribosylformylgl
F;4-317/Domain: phosphoribosylformylglycinamidine cyclo-ligase homology <PFCL>

  Query Match             51.4%;  Score 38;  DB 2;  Length 331;
  Best Local Similarity   63.6%;  Pred. No. 23;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 ITYRISGVGID 15
          :||| :|| ||
Db      2 VTYRSAGVDID 12


Search completed: January  7, 2000, 08:52:37
Job time: 2267 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:19:08 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-1
Perfect score: 74
Sequence: 1 ATQKITYRISGVGID 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	999	1	DSG3_HUMAN	P32926 homo sapien
2	64	86.5	1043	1	DSG1_BOVIN	Q03763 bos taurus
3	64	86.5	1049	1	DSG1_HUMAN	Q02413 homo sapien
4	44	59.5	388	1	PUR5_VIGUN	P52424 vigna ungui
5	42	56.8	897	1	BGAL_CLOAB	P24131 clostridium
6	42	56.8	550	1	CBS_HUMAN	P35520 homo sapien
7	39	52.7	504	1	MYOC_HUMAN	Q99972 homo sapien
8	38	51.4	140	1	HMGL_BOVIN	Q29448 bos taurus
9	38	51.4	298	1	HMGL_CHICK	P35915 gallus gall
10	38	51.4	325	1	HMGL_MOUSE	P38060 mus musculu
11	38	51.4	325	1	HMGL_RAT	P97519 rattus norv
12	38	51.4	956	1	TSP3_MOUSE	Q05895 mus musculu
13	37	50.0	560	1	CBS_RAT	P32232 rattus norv
14	37	50.0	1117	1	DSG2_HUMAN	Q14126 homo sapien
15	37	50.0	325	1	HMGL_HUMAN	P35914 homo sapien
16	37	50.0	1363	1	VGL2_CVBF	P25190 bovine coro
17	37	50.0	1363	1	VGL2_CVBL9	P25191 bovine coro
18	37	50.0	1363	1	VGL2_CVBLY	P25192 bovine coro
19	37	50.0	1363	1	VGL2_CVBM	P15777 bovine coro
20	37	50.0	1363	1	VGL2_CVBQ	P25193 bovine coro
21	37	50.0	1363	1	VGL2_CVBV	P25194 bovine coro
22	37	50.0	1121	1	YJ09_YEAST	P47107 saccharomyc
23	37	50.0	1218	1	YK83_YEAST	P36028 saccharomyc
24	36	48.6	890	1	BCN5_CLOPE	P08696 clostridium
25	36	48.6	712	1	CADD_CHICK	P33150 gallus gall
26	36	48.6	761	1	METE_BACSU	P80877 b 5-methylt
27	36	48.6	690	1	NCPR_PHAAU	P37116 phaseolus a
28	36	48.6	247	1	OPAG_NEIGO	Q04875 neisseria g
29	36	48.6	740	1	PFL_CLOPA	Q46266 clostridium
30	36	48.6	355	1	PUR5_ARATH	Q05728 arabidopsis
31	36	48.6	323	1	YF78_HAEIN	Q57287 haemophilus
32	35	47.3	420	1	A1AU_HUMAN	P20848 homo sapien
33	35	47.3	859	1	ABR_HUMAN	Q12979 homo sapien
34	35	47.3	985	1	AGLU_ASPNG	P56526 aspergillus
35	35	47.3	127	1	C560_CHOCR	P48934 chondrus cr
36	35	47.3	882	1	CAD1_HUMAN	P12830 homo sapien
37	35	47.3	884	1	CAD1_MOUSE	P09803 mus musculu
38	35	47.3	913	1	CAD4_CHICK	P24503 gallus gall
39	35	47.3	916	1	CAD4_HUMAN	P55283 homo sapien
40	35	47.3	913	1	CAD4_MOUSE	P39038 mus musculu
41	35	47.3	509	1	CRU3_BRANA	P33525 brassica na
42	35	47.3	893	1	DSC1_BOVIN	Q01107 bos taurus
43	35	47.3	894	1	DSC1_HUMAN	Q08554 homo sapien
44	35	47.3	863	1	DSC2_BOVIN	P33545 bos taurus
45	35	47.3	607	1	G6PI_TRYBB	P13377 trypanosoma

ALIGNMENTS

RESULT 1
DSG3_HUMAN
ID DSG3_HUMAN STANDARD; PRT; 999 AA.
AC P32926;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92069753.
RA AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion.";
RL Cell 67:869-877(1991).
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC ----------

CC ----------
DR EMBL; M76482; AAA60230.1; -.
DR PIR; A41088; IJHUG3.
DR HSSP; P09803; 1EDH.
DR MIM; 169615; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.

FT	SIGNAL	1	23 POTENTIAL.
FT	PROPEP	24	49 POTENTIAL.
FT	CHAIN	50	999 DESMOGLEIN 3.
FT	DOMAIN	50	615 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	616	640 POTENTIAL.
FT	DOMAIN	641	999 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50	158 CADHERIN 1.
FT	REPEAT	159	268 CADHERIN 2.
FT	REPEAT	269	383 CADHERIN 3.
FT	REPEAT	386	499 CADHERIN 4.
FT	REPEAT	910	935 DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966 DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110	110 POTENTIAL.

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match             100.0%;  Score 74;  DB 1;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          |||||||||||||||
Db     79 ATQKITYRISGVGID 93


RESULT   2
DSG1_BOVIN
ID   DSG1_BOVIN     STANDARD;      PRT;  1043 AA.
AC   Q03763;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
GN   DSG1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANKE W.W.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 44-1043 FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 91168965.
RA   KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA   ZIMBELMANN R., FRANKE W.W.;
RT   "Identification of desmoglein, a constitutive desmosomal
RT   glycoprotein, as a member of the cadherin family of cell adhesion
RT   molecules.";
RL   Eur. J. Cell Biol. 53:1-12(1990).
RN   [3]
RP   REVISIONS, AND SEQUENCE OF 101-123.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA   FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
RN   [4]
RP   SEQUENCE OF 44-493 FROM N.A.
RX   MEDLINE; 91097553.
RA   GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT   "Desmoglein shows extensive homology to the cadherin family of cell
RT   adhesion molecules.";
RL   Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X58466; CAA41380.1; -.
DR   EMBL; X57784; CAA40930.1; -.
DR   EMBL; M58165; AAA62709.1; -.
DR   PIR; S14603; IJBOG1.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1043       DESMOGLEIN 1.
FT   DOMAIN       50    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    549    573       POTENTIAL.
FT   DOMAIN      574   1043       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    498       CADHERIN 4.
FT   REPEAT      819    845       DESMOGLEIN REPEAT 1.
FT   REPEAT      846    875       DESMOGLEIN REPEAT 2.
FT   REPEAT      876    905       DESMOGLEIN REPEAT 3.
FT   REPEAT      906    933       DESMOGLEIN REPEAT 4.
FT   REPEAT      934    962       DESMOGLEIN REPEAT 5.
FT   DOMAIN      963   1012       GLY/SER-RICH.
FT   CARBOHYD    110    110
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    496    496       POTENTIAL.
FT   CONFLICT    124    124       I -> V (IN REF. 4).
SQ   SEQUENCE   1043 AA;  112243 MW;  13898584 CRC32;


  Query Match             86.5%;  Score 64;  DB 1;  Length 1043;
  Best Local Similarity   80.0%;  Pred. No. 0.00098;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          | |::|||||||||||
Db     79 ANQQVTYRISGVGID 93


RESULT   3
DSG1_HUMAN
ID   DSG1_HUMAN     STANDARD;      PRT;  1049 AA.
AC   Q02413;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN   DSG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 91271279.
RA   WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOTIS P., POYNTER D.,
RA   ARNEMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA   BUXTON R.S., MAGEE A.I.;
RT   "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT   junctions, is related to the cadherin family of cell adhesion
RT   molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
```

```
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56654; CAA39976.1; -.
DR   PIR; S16906; IJHUG1.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125670; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1049       DESMOGLEIN 1.
FT   DOMAIN       50    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    570       POTENTIAL.
FT   DOMAIN      571   1049       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    497       CADHERIN 4.
FT   REPEAT      813    839       DESMOGLEIN REPEAT 1.
FT   REPEAT      840    869       DESMOGLEIN REPEAT 2.
FT   REPEAT      870    899       DESMOGLEIN REPEAT 3.
FT   REPEAT      900    927       DESMOGLEIN REPEAT 4.
FT   REPEAT      928    956       DESMOGLEIN REPEAT 5.
FT   DOMAIN      969   1019       GLY/SER-RICH.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    180    180       POTENTIAL.
SQ   SEQUENCE   1049 AA;  113715 MW;  FDD79961 CRC32;


  Query Match             86.5%;  Score 64;  DB 1;  Length 1049;
  Best Local Similarity   80.0%;  Pred. No. 0.00099;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          | |::|||||||||||
Db     79 ANQQVTYRISGVGID 93


RESULT   4
PUR5_VIGUN
ID   PUR5_VIGUN     STANDARD;      PRT;   388 AA.
AC   P52424;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE PRECURSOR (EC 6.3.3.1)
DE   (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR SYNTHASE).
GN   PUR5.
OS   Vigna unguiculata (Cowpea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC   Vigna.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VITA 3; TISSUE=ROOT NODULES;
RA   SMITH P.M.C., MANN A.J., HALL D.J., ATKINS C.A.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE =
CC       ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-5-AMINOIMIDAZOLE.
CC   -!- PATHWAY: FIFTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- SIMILARITY: TO OTHER AIRS FROM BACTERIA AND EUKARYOTES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30895; AAC14578.1; -.
DR   PFAM; PF00586; AIRS; 1.
KW   Purine biosynthesis; Ligase; Chloroplast; Transit peptide.
FT   TRANSIT       1      ?       CHLOROPLAST (POTENTIAL).
FT   CHAIN         ?    388       PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-
FT                                LIGASE.
SQ   SEQUENCE   388 AA;  40427 MW;  895D3159 CRC32;


  Query Match             59.5%;  Score 44;  DB 1;  Length 388;
  Best Local Similarity   60.0%;  Pred. No. 1.4;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
            |:| :||| :|| ||
Db     62 ASQGLTYRDAGVDID 76


RESULT   5
BGAL_CLOAB
ID   BGAL_CLOAB     STANDARD;      PRT;   897 AA.
AC   P24131;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN   CBGA.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 2951;
RX   MEDLINE; 91216979.
RA   HANCOCK K.R., ROCKMAN E., YOUNG C.A., PEARCE L., MADDOX I.S.,
RA   SCOTT D.B.;
RT   "Expression and nucleotide sequence of the Clostridium acetobutylicum
RT   beta-galactosidase gene cloned in Escherichia coli.";
RL   J. Bacteriol. 173:3084-3095(1991).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC       GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC   -!- INDUCTION: LATE IN THE ABE (ACETONE, BUTANOL, AND ETHANOL)
CC       FERMENTATION AND SUBJECT TO GLUCOSE REPRESSION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35107; AAA23216.1; -.
DR   PIR; A39405; A39405.
DR   HSSP; P00722; 1BGL.
DR   PFAM; PF00703; Glyco_hydro_2; 1.
DR   PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR   PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW   Hydrolase; Glycosidase.
FT   ACT_SITE    459    459       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    503    503       NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   897 AA;  105020 MW;  5CFF268B CRC32;
```

```
  Query Match             56.8%;  Score 42;  DB 1;  Length 897;
  Best Local Similarity   72.7%;  Pred. No. 7.5;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 KITYRISGVGI 14
          |||| :|| ||
Db    828 KITYEVSGEGI 838


RESULT   6
CBS_HUMAN
ID   CBS_HUMAN      STANDARD;      PRT;   550 AA.
AC   P35520; Q99425;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYSTATHIONINE BETA-SYNTHASE (EC 4.2.1.22) (SERINE SULFHYDRASE)
DE   (BETA-THIONASE).
GN   CBS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 94093551.
RA   KRAUS J.P., LE K., SWAROOP M., OHURA T., TAHARA T., ROSENBERG L.E.,
RA   ROPER M.D., KOZICH V.;
RT   "Human cystathionine beta-synthase cDNA: sequence, alternative
RT   splicing and expression in cultured cells.";
RL   Hum. Mol. Genet. 2:1633-1638(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95321944.
RA   CHASSE J.-F., PALY E., PARIS D., PAUL V., SINET P.M., KAMOUN P.,
RA   LONDON J.;
RT   "Genomic organization of the human cystathionine beta-synthase gene:
RT   evidence for various cDNAs.";
RL   Biochem. Biophys. Res. Commun. 211:826-832(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94294429.
RA   KRUGER W.D., COX D.R.;
RT   "A yeast system for expression of human cystathionine beta-synthase:
RT   structural and functional conservation of the human and yeast genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6614-6618(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98051926.
RA   CHASSE J.F., PAUL V., ESCANEZ R., KAMOUN P., LONDON J.;
RT   "Human cystathionine beta-synthase: gene organization and expression
RT   of different 5' alternative splicing.";
RL   Mamm. Genome 8:917-921(1997).
RN   [5]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE; 99009316.
RA   KRAUS J.P., OLIVERIUSOVA J., SOKOLOVA J., KRAUS E., VLCEK C.,
RA   DE FRANCHIS R., MACLEAN K.N., BAO L., BUKOVSKA G., PATTERSON D.,
RA   PACES V., ANSORGE W., KOZICH V.;
RT   "The human cystathionine beta-synthase (CBS) gene: complete sequence,
RT   alternative splicing, and polymorphisms.";
RL   Genomics 52:312-324(1998).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE; 78242360.
RA   KRAUS J.P., PACKMAN S., FOWLER B., ROSENBERG L.E.;
RT   "Purification and properties of cystathionine beta-synthase from
RT   human liver. Evidence for identical subunits.";
RL   J. Biol. Chem. 253:6523-6528(1978).
RN   [7]
RP   VARIANT THR-277.
RX   MEDLINE; 93244841.
RA   KOZICH V., KRAUS J.P.;
RT   "Screening for mutations by expressing patient cDNA segments in E.
RT   coli: homocystinuria due to cystathionine beta-synthase deficiency.";
RL   Hum. Mutat. 1:113-123(1992).
RN   [8]
RP   VARIANTS VAL-113 AND LEU-144.
RX   MEDLINE; 93357760.
RA   KOZICH V., DE FRANCHIS R., KRAUS J.P.;
RT   "Molecular defect in a patient with pyridoxine-responsive
RT   homocystinuria.";
RL   Hum. Mol. Genet. 2:815-816(1993).
RN   [9]
RP   VARIANTS THR-277 AND SER-306.
RX   MEDLINE; 94108438.
RA   HU F.L., GU Z., KOZICH V., KRAUS J.P., RAMESH V., SHIH V.E.;
RT   "Molecular basis of cystathionine beta-synthase deficiency in
RT   pyridoxine responsive and nonresponsive homocystinuria.";
RL   Hum. Mol. Genet. 2:1857-1860(1993).
RN   [10]
RP   VARIANTS ARG-77 AND ASN-101.
RX   MEDLINE; 95072574.
RA   DE FRANCHIS R., KOZICH V., MCINNES R., KRAUS J.P.;
RT   "Identical genotypes in siblings with different homocystinuric
RT   phenotypes: identification of three mutations in cystathionine beta-
RT   synthase using an improved bacterial expression system.";
RL   Hum. Mol. Genet. 3:1103-1108(1994).
RN   [11]
RP   VARIANTS GLN-124 AND ASP-130.
RX   MEDLINE; 95152517.
RA   MARBLE M., GERAGHTY M.T., DE FRANCHIS R., KRAUS J.P., VALLE D.;
RT   "Characterization of a cystathionine beta-synthase allele with three
RT   mutations in cis in a patient with B6 nonresponsive homocystinuria.";
RL   Hum. Mol. Genet. 3:1883-1886(1994).
RN   [12]
RP   VARIANTS.
RX   MEDLINE; 95056993.
RA   KRAUS J.P.;
RT   "Komrower Lecture. Molecular basis of phenotype expression in
RT   homocystinuria.";
RL   J. Inherit. Metab. Dis. 17:383-390(1994).
RN   [13]
RP   VARIANTS SER-87; GLN-124 AND MET-256.
RX   MEDLINE; 95282779.
RA   SEBASTIO G., SPERANDEO M.P., PANICO M., DE FRANCHIS R., KRAUS J.P.,
RA   ANDRIA G.;
RT   "The molecular basis of homocystinuria due to cystathionine beta-
RT   synthase deficiency in Italian families, and report of four novel
RT   mutations.";
RL   Am. J. Hum. Genet. 56:1324-1333(1995).
RN   [14]
RP   VARIANTS TYR-164 AND MET-370.
RX   MEDLINE; 95362263.
RA   KLUIJTMANS L.A.J., BLOM H.J., BOERS G.H.J., VAN OOST B.A.,
RA   TRIJBELS F.J.M., VAN DEN HEUVEL L.P.W.J.;
RT   "Two novel missense mutations in the cystathionine beta-synthase gene
RT   in homocystinuric patients.";
RL   Hum. Genet. 96:249-250(1995).
RN   [15]
RP   VARIANTS MET-167; HIS-223; THR-277; GLY-306; VAL-330 AND GLU-453.
RX   MEDLINE; 96133289.
RA   KRUGER W.D., COX D.R.;
RT   "A yeast assay for functional detection of mutations in the human
RT   cystathionine beta-synthase gene.";
RL   Hum. Mol. Genet. 4:1155-1161(1995).
RN   [16]
RP   VARIANT LEU-289.
RX   MEDLINE; 96057367.
RA   SPERANDEO M.P., PANICO M., PEPE A., CANDITO M., DE FRANCHIS R.,
RA   KRAUS J.P., ANDRIA G., SEBASTIO G.;
RT   "Molecular analysis of patients affected by homocystinuria due to
RT   cystathionine beta-synthase deficiency: report of a new mutation in
```

```
RT   exon 8 and a deletion in intron 11.";
RL   J. Inherit. Metab. Dis. 18:211-214(1995).
RN   [17]
RP   VARIANTS GLU-383 AND SER-538.
RX   MEDLINE; 97144192.
RA   ARAL B., COUDE M., LONDON J., AUPETIT J., CHASSE J.-F., ZABOT M.-T.,
RA   CHADEFAUX-VEKEMANS B., KAMOUN P.;
RT   "Two novel mutations (K384E and L539S) in the C-terminal moiety of the
RT   cystathionine beta-synthase protein in two French pyridoxine-
RT   responsive homocystinuria patients.";
RL   Hum. Mutat. 9:81-82(1997).
RN   [18]
RP   VARIANT ASN-443.
RX   MEDLINE; 96331231.
RA   KLUIJTMANS L.A.J., BOERS G.H.J., STEVENS E.M.B., RENIER W.O.,
RA   KRAUS J.P., TRIJBELS F.J.M., VAN DEN HEUVEL L.P.W.J., BLOM H.J.;
RT   "Defective cystathionine beta-synthase regulation by S-
RT   adenosylmethionine in a partially pyridoxine responsive homocystinuria
RT   patient.";
RL   J. Clin. Invest. 98:285-289(1996).
RN   [19]
RP   VARIANTS LYS-143 AND TYR-164.
RA   GORDON R.B., COX A.J., DAWSON P.A., EMMERSON B.T., KRAUS J.P.,
RA   DUDMAN N.P.B.;
RT   "Mutational analysis of the cystathionine beta-synthase gene: a
RT   splicing mutation, two missense mutations and an insertion in patients
RT   with homocystinuria.";
RL   Hum. Mutat. 11:332-332(1998).
CC   -!- CATALYTIC ACTIVITY: L-SERINE + L-HOMOCYSTEINE = CYSTATHIONINE +
CC       H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: FIRST STEP IN HOMOCYSTEINE TRANSULFURATION.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS; A MAJOR (SHOWN HERE) AND A
CC       MINOR ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN THE ADULT STRONGLY EXPRESSED IN LIVER AND
CC       PANCREAS, SOME EXPRESSION IN HEART AND BRAIN, WEAK EXPRESSION IN
CC       LUNG AND KIDNEY. IN THE FETUS, EXPRESSED IN BRAIN, LIVER AND
CC       KIDNEY.
CC   -!- DISEASE: DEFECTS IN CBS ARE A MAJOR CAUSE OF HOMOCYSTINURIA. THE
CC       MOST FREQUENT SYMPTOMS INCLUDE DISLOCATED OPTIC LENSES, VASCULAR
CC       DISORDERS (ARTERIOSCLEROSIS AND THROMBOSIS), SKELETAL
CC       ABNORMALITIES AND MENTAL RETARDATION. PATIENTS WITH HOMOCYSTENURIA
CC       HAVE ELEVATED LEVELS OF HOMOCYSTEINE AND METHIONINE IN THEIR BODY
CC       FLUID.
CC   -!- SIMILARITY: BELONGS TO THE CYSTEINE SYNTHASE/CYSTATHIONINE BETA-
CC       SYNTHASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CBS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L19501; AAA19874.1; -.
DR   EMBL; X82166; CAA57656.1; -.
DR   EMBL; L14577; AAA98524.1; -.
DR   EMBL; X88562; CAA61252.1; -.
DR   EMBL; X91910; CAA61252.1; JOINED.
DR   EMBL; X98811; CAA61252.1; JOINED.
DR   EMBL; X98812; CAA61252.1; JOINED.
DR   EMBL; X98813; CAA61252.1; JOINED.
DR   EMBL; X98814; CAA61252.1; JOINED.

  Query Match           56.8%;  Score 42;  DB 1;  Length 550;
  Best Local Similarity  42.9%;  Pred. No. 4.6;
  Matches    6;  Conservative    4;  Mismatches   4;  Indels    0;  Gaps    0;

Qy       2 TQKITYRISGVGID 15
           |:: || : |:| |
Db     295 TEQTTYEVEGIGYD 308


RESULT   7
MYOC_HUMAN
ID   MYOC_HUMAN     STANDARD;      PRT;   504 AA.
AC   Q99972; O00620;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MYOCILIN PRECURSOR (TRABECULAR MESHWORK-INDUCED GLUCOCORTICOID
DE   RESPONSE PROTEIN).
GN   MYOC OR TIGR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS JOAG VAL-364 AND HIS-437.
RX   MEDLINE; 97158493.
RA   STONE E.M., FINGERT J.H., ALWARD W.L.M., NGUYEN T.D., POLANSKY J.R.,
RA   SUNDEN S.L.F., NISHIMURA D., CLARK A.F., NYSTUEN A., NICHOLS B.E.,
RA   MACKEY D.A., RITCH R., KALENAK J.W., CRAVEN E.R., SHEFFIELD V.C.;
RT   "Identification of a gene that causes primary open angle glaucoma.";
RL   Science 275:668-670(1997).
RN   [2]
RP   REVISIONS, AND SEQUENCE OF 1-6 AND 33-37.
RX   MEDLINE; 98165818.
RA   NGUYEN T.D., CHEN P., HUANG W.D., CHEN H., JOHNSON D., POLANSKY J.R.;
RT   "Gene structure and properties of TIGR, an olfactomedin-related
RT   glycoprotein cloned from glucocorticoid-induced trabecular meshwork
RT   cells.";
RL   J. Biol. Chem. 273:6341-6350(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97424389.
RA   ORTEGO J., ESCRIBANO J., COCA-PRADOS M.;
RT   "Cloning and characterization of subtracted cDNAs from a human
RT   ciliary body library encoding TIGR, a protein involved in juvenile
RT   open angle glaucoma with homology to myosin and olfactomedin.";
RL   FEBS Lett. 413:349-353(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 97312692.
RA   KUBOTA R., NODA S., WANG Y., MINOSHIMA S., ASAKAWA S., KUDOH J.,
RA   MASHIMA Y., OGUCHI Y., SHIMIZU N.;
RT   "A novel myosin-like protein (myocilin) expressed in the connecting
RT   cilium of the photoreceptor: molecular cloning, tissue expression,
RT   and chromosomal mapping.";
RL   Genomics 41:360-369(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
RA   GARCHON H.-J.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   DEADMAN R.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98217378.
RA   FINGERT J.H., YING L., SWIDERSKI R.E., NYSTUEN A.M., ARBOUR N.C.,
RA   ALWARD W.L.M., SHEFFIELD V.C., STONE E.M.;
RT   "Characterization and comparison of the human and mouse GLC1A glaucoma
RT   genes.";
RL   Genome Res. 8:377-384(1998).
RN   [8]
RP   VARIANTS JOAG ARG-367 AND LEU-370.
RX   MEDLINE; 98008006.
RA   SUZUKI Y., SHIRATO S., TANIGUCHI F., OHARA K., NISHIMAKI K., OHTA S.;
RT   "Mutations in the TIGR gene in familial primary open-angle glaucoma
```

```
RT   in Japan.";
RL   Am. J. Hum. Genet. 61:1202-1204(1997).
RN   [9]
RP   VARIANTS POAG ARG-246; LEU-370; SER-477; LYS-480 AND PHE-499.
RX   MEDLINE; 97472461.
RA   ADAM M.F., BELMOUDEN A., BINISTI P., BREZIN A.P., VALTOT F.,
RA   BECHETOILLE A., DASCOTTE J.-C., COPIN B., GOMEZ L., CHAVENTRE A.,
RA   BACH J.-F., GARCHON H.-J.;
RT   "Recurrent mutations in a single exon encoding the evolutionarily
RT   conserved olfactomedin-homology domain of TIGR in familial open-angle
RT   glaucoma.";
RL   Hum. Mol. Genet. 6:2091-2097(1997).
RN   [10]
RP   VARIANT JOAG ARG-337.
RX   MEDLINE; 98027214.
RA   STOILOVA D., CHILD A., BRICE G., CRICK R.P., FLECK B.W., SARFARAZI M.;
RT   "Identification of a new 'TIGR' mutation in a family with juvenile-
RT   onset primary open angle glaucoma.";
RL   Ophthalmic Genet. 18:109-118(1997).
RN   [11]
RP   VARIANTS JOAG ARG-367 AND LEU-370.
RX   MEDLINE; 98141135.
RA   MICHELS-RAUTENSTRAUSS K.G., MARDIN C.Y., BUDDE W.M., LIEHR T.,
RA   POLANSKY J., NGUYEN T., TIMMERMAN V., VAN BROECKHOVEN C.,
RA   NAUMANN G.O.H., PFEIFFER R.A., RAUTENSTRAUSS B.W.;
RT   "Juvenile open angle glaucoma: fine mapping of the TIGR gene to
RT   1q24.3-q25.2 and mutation analysis.";
RL   Hum. Genet. 102:103-106(1998).
RN   [12]
RP   VARIANT COAG ARG-367, AND VARIANT JOAG PHE-426.
RX   MEDLINE; 98180724.
RA   MANSERGH F.C., KENNA P.F., AYUSO C., KIANG A.-S., HUMPHRIES P.,
RA   FARRAR G.J.;
RT   "Novel mutations in the TIGR gene in early and late onset open angle
RT   glaucoma.";
RL   Hum. Mutat. 11:244-251(1998).
CC   -!- FUNCTION: MAY PARTICIPATE IN THE OBSTRUCTION OF FLUID OUTFLOW IN
CC       THE TRABECULAR MESHWORK.
CC   -!- SUBCELLULAR LOCATION: LOCATED PREFERENTIALLY IN THE CILIARY
CC       ROOTLET AND BASAL BODY OF THE CONNECTING CILIUM OF PHOTORECEPTOR
CC       CELLS, AND IN THE ROUGH ENDOPLASMIC RETICULUM. ALSO SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: THE SHORT FORM MAY BE THE PRODUCT OF AN
CC       ALTERNATIVE INITIATION AT MET-15.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LARGE AMOUNTS IN VARIOUS TYPES OF
CC       MUSCLE, CILIARY BODY, PAPILLARY SPHINCTER, SKELETAL MUSCLE, HEART
CC       AND OTHER TISSUES. EXPRESSED PREDOMINANTLY IN THE RETINA. IN
CC       NORMAL EYES, FOUND IN THE INNER UVEAL MESHWORK REGION AND THE
CC       ANTERIOR PORTION OF THE MESHWORK. IN CONTRAST, IN MANY
CC       GLAUCOMATOUS EYES, IT IS FOUND IN MORE REGIONS OF THE MESHWORK AND
CC       APPEARED MORE INTENSIVELY THAN IN NORMAL EYES, REGARDLESS OF THE
CC       TYPE OR CLINICAL SEVERITY OF GLAUCOMA.
CC   -!- PTM: DIFFERENT ISOFORMS MAY ARISE BY POST-TRANSLATIONAL
CC       MODIFICATIONS.
CC   -!- DISEASE: DEFECTS IN MYOC ARE A CAUSE OF THE PRIMARY OPEN ANGLE
CC       GLAUCOMA (POAG). POAG IS SUBDIVIDED INTO TWO DIFFERENT CATEGORIES
CC       DEPENDING ON THE AGE OF ONSET: JUVENILE ONSET OPEN ANGLE GLAUCOMA
CC       (JOAG OR GLC1A) AND CHRONIC OPEN ANGLE GLAUCOMA (COAG) WHICH IS
CC       LATER IN ONSET. POAG IS A HIGHLY PREVALENT CAUSE OF IRREVERSIBLE
CC       BLINDNESS. IT IS CHARACTERIZED BY CUPPING OF THE OPTIC DISK AND
CC       ALTERATION OF THE VISUAL FIELD. ELEVATION OF INTRAOCULAR PRESSURE
CC       IS OFTEN PRESENT AND IS A MAJOR RISK FACTOR. THE DISEASE IS
CC       PAINLESS AND OFTEN DIAGNOSED AT A LATE STAGE, WHEN VISUAL FIELD
CC       DEFECTS ARE SEVERE.
CC   -!- SIMILARITY: BELONGS TO THE OLFACTOMEDIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U85257; AAC52051.1; -.
DR   EMBL; AF001620; AAC51725.1; -.
DR   EMBL; D88214; BAA23531.1; -.
DR   EMBL; Z97171; CAB09899.1; -.
DR   EMBL; Z97177; CAB09899.1; JOINED.
DR   EMBL; Z97174; CAB09899.1; JOINED.
DR   EMBL; Z98750; CAB11430.1; -.
DR   EMBL; AF049793; AAC14264.1; -.
DR   EMBL; AF049791; AAC14264.1; JOINED.
DR   EMBL; AF049792; AAC14264.1; JOINED.
DR   HSSP; P01100; 1FOS.
DR   MIM; 601652; -.
DR   MIM; 137750; -.
KW   Signal; Disease mutation; Glycoprotein; Alternative initiation;
KW   Coiled coil.
FT   SIGNAL       1    32       POTENTIAL.
FT   CHAIN       33   504       MYOCILIN.
FT   DOMAIN      74   184       COILED COIL (POTENTIAL).
FT   DOMAIN     310   486       OLFACTOMEDIN-LIKE.
FT   CARBOHYD    57    57       POTENTIAL.
FT   SITE       502   504       MICROBODY TARGETING SIGNAL (POTENTIAL).
FT   VARIANT    246   246       G -> R (IN JOAG).
FT                              /FTId=VAR_005468.
FT   VARIANT    337   337       Q -> R (IN JOAG).
FT                              /FTId=VAR_005469.
FT   VARIANT    364   364       G -> V (IN JOAG).
FT                              /FTId=VAR_005470.
FT   VARIANT    367   367       G -> R (IN COAG).
FT                              /FTId=VAR_005471.
FT   VARIANT    370   370       P -> L (IN JOAG).
FT                              /FTId=VAR_005472.
FT   VARIANT    426   426       V -> F (IN JOAG).
FT                              /FTId=VAR_005473.
FT   VARIANT    437   437       Y -> H (IN JOAG).
FT                              /FTId=VAR_005474.
FT   VARIANT    477   477       I -> S (IN JOAG AND COAG).
FT                              /FTId=VAR_005475.
FT   VARIANT    480   480       N -> K (IN JOAG AND COAG).
FT                              /FTId=VAR_005476.
FT   VARIANT    499   499       I -> F (IN COAG).
FT                              /FTId=VAR_005477.
SQ   SEQUENCE   504 AA;  56972 MW;  1919A632 CRC32;


  Query Match             52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 14;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 TQKITYRISGVGID 15
           ||: |:|| || | |
Db     281 TQETTWRIDTVGTD 294


RESULT   8
HMGL_BOVIN
ID   HMGL_BOVIN     STANDARD;      PRT;   140 AA.
AC   Q29448;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL) (3-
DE   HYDROXY-3-METHYLGLUTARATE-COA LYASE) (FRAGMENT).
GN   HMGCL.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   JI S., KUSKE J., SPURLOCK M.E.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA =
CC       ACETYL-COA + ACETOACETATE.
```

```
CC   -!- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41409; AAA86757.1; -.
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase; Mitochondrion.
FT   NON_TER       1      1
FT   ACT_SITE     81     81       BY SIMILARITY.
SQ   SEQUENCE   140 AA;  14409 MW;  08C3CDC3 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 140;
  Best Local Similarity   42.9%;  Pred. No. 6.1;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db     92 ATEDLVYMLAGLGI 105


RESULT   9
HMGL_CHICK
ID   HMGL_CHICK     STANDARD;      PRT;   298 AA.
AC   P35915;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL) (3-
DE   HYDROXY-3-METHYLGLUTARATE-COA LYASE).
GN   HMGCL.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE OF 10-298 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 93179448.
RA   MITCHELL G.A., ROBERT M.-F., HRUZ P.W., WANG S., FONTAINE G.,
RA   BEHNKE C.E., MENDE-MUELLER L.M., SCHAPPERT K., LEE C., GIBSON K.M.,
RA   MIZIORKO H.M.;
RT   "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL). Cloning of human
RT   and chicken liver HL cDNAs and characterization of a mutation causing
RT   human HL deficiency.";
RL   J. Biol. Chem. 268:4376-4381(1993).
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA =
CC       ACETYL-COA + ACETOACETATE.
CC   -!- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07034; AAA92728.1; ALT_INIT.
DR   PIR; B45470; B45470.
DR   PFAM; PF00682; HMGL-like; 1.
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase; Mitochondrion.
FT   ACT_SITE    239    239       BY SIMILARITY.
SQ   SEQUENCE   298 AA;  31436 MW;  A2EE0702 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 298;
  Best Local Similarity   42.9%;  Pred. No. 13;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db    250 ATEDLVYMLNGLGI 263


RESULT  10
HMGL_MOUSE
ID   HMGL_MOUSE     STANDARD;      PRT;   325 AA.
AC   P38060;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)
DE   (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLUTARATE-COA LYASE).
GN   HMGCL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93364116.
RA   WANG S., NADEAU J.H., DUNCAN A., ROBERT M.-F., FONTAINE G.,
RA   SCHAPPERT K., JOHNSON K.R., ZIETKIEWICZ E., HRUZ P., MIZIORKO H.;
RT   "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL): cloning and
RT   characterization of a mouse liver HL cDNA and subchromosomal mapping
RT   of the human and mouse HL genes.";
RL   Mamm. Genome 4:382-387(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE; 96207305.
RA   WANG S.P., ROBERT M.-F., GIBSON K.M., WANDERS R.J.A., MITCHELL G.A.;
RT   "3-hydroxy-3-methylglutaryl CoA lyase (HL): mouse and human HL gene
RT   (HMGCL) cloning and detection of large gene deletions in two
RT   unrelated HL-deficient patients.";
RL   Genomics 33:99-104(1996).
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA =
CC       ACETYL-COA + ACETOACETATE.
CC   -!- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S65036; AAB27965.1; -.
DR   EMBL; U49878; AAB03107.1; -.
DR   EMBL; U49870; AAB03107.1; JOINED.
DR   EMBL; U49871; AAB03107.1; JOINED.
DR   EMBL; U49872; AAB03107.1; JOINED.
DR   EMBL; U49873; AAB03107.1; JOINED.
DR   EMBL; U49874; AAB03107.1; JOINED.
DR   EMBL; U49875; AAB03107.1; JOINED.
DR   EMBL; U49876; AAB03107.1; JOINED.
DR   EMBL; U49877; AAB03107.1; JOINED.
DR   MGD; MGI:96158; HMGCL.
DR   PFAM; PF00682; HMGL-like; 1.
```

```
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase; Mitochondrion; Transit peptide.
FT   TRANSIT       1     27       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN        28    325       HYDROXYMETHYLGLUTARYL-COA LYASE.
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   CONFLICT     62     63       ML -> IV (IN REF. 2).
FT   CONFLICT     80     80       N -> K (IN REF. 2).
FT   CONFLICT    231    231       G -> A (IN REF. 2).
FT   CONFLICT    238    238       I -> Y (IN REF. 2).
SQ   SEQUENCE   325 AA;  34161 MW;  3448C502 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 325;
  Best Local Similarity   42.9%;  Pred. No. 14;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db    277 ATEDLVYMLNGLGI 290


RESULT  11
HMGL_RAT
ID   HMGL_RAT       STANDARD;      PRT;   325 AA.
AC   P97519;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)
DE   (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLUTARATE-COA LYASE).
GN   HMGCL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RA   CULLINGFORD T.E., DOLPHIN C.T., CANEVARI L., CLARK J.B.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA =
CC       ACETYL-COA + ACETOACETATE.
CC   -!- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y10054; CAA71148.1; -.
DR   PFAM; PF00682; HMGL-like; 1.
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase; Mitochondrion; Transit peptide.
FT   TRANSIT       1     27       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN        28    325       HYDROXYMETHYLGLUTARYL-COA LYASE.
FT   ACT_SITE    266    266       BY SIMILARITY.
SQ   SEQUENCE   325 AA;  34192 MW;  C68DE719 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 325;
  Best Local Similarity   42.9%;  Pred. No. 14;
  Matches    6;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:||
Db    277 ATEDLVYMLTGLGI 290


RESULT  12
TSP3_MOUSE
ID   TSP3_MOUSE     STANDARD;      PRT;   956 AA.
AC   Q05895;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   THROMBOSPONDIN 3 PRECURSOR.
GN   THBS3 OR TSP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBA; TISSUE=LUNG;
RX   MEDLINE; 93224149.
RA   BORNSTEIN P., DEVARAYALU S., EDELHOFF S., DISTECHE C.M.;
RT   "Isolation and characterization of the mouse thrombospondin 3 (Thbs3)
RT   gene.";
RL   Genomics 15:607-613(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94117438.
RA   QABAR A.N., LIN Z., WOLF F.W., O'SHEA K.S., LAWLER J., DIXIT V.M.;
RT   "Thrombospondin 3 is a developmentally regulated heparin binding
RT   protein.";
RL   J. Biol. Chem. 269:1262-1269(1994).
RN   [3]
RP   SEQUENCE OF 517-956 FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE; 92291102.
RA   VOS H.L., DEVARAYALU S., DE VRIES Y., BORNSTEIN P.;
RT   "Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene
RT   family.";
RL   J. Biol. Chem. 267:12192-12196(1992).
RN   [4]
RP   SEQUENCE OF 834-956 FROM N.A.
RC   STRAIN=BALB/C;
RA   VOS H.L., MOCKINGSTURM-WILSON M., ROOD P.M.L., MAAS A.C.E.,
RA   DUHIG T., GENDLER S.J., BORNSTEIN P.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ AND TYPE V COLLAGEN.
CC   -!- SUBUNIT: OLIGOMER, CROSS-LINKED BY DISULFIDE BONDS.
CC   -!- TISSUE SPECIFICITY: BRAIN, LUNGS AND CARTILAGE.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TYPE-3 TSP REPEATS (WHICH BIND CALCIUM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04302; AAA40497.1; -.
DR   EMBL; L24434; AAA40433.1; -.
DR   EMBL; M86620; AAA40430.1; -.
DR   EMBL; M86611; AAA40430.1; JOINED.
DR   EMBL; M86612; AAA40430.1; JOINED.
DR   EMBL; M86613; AAA40430.1; JOINED.
DR   EMBL; M86614; AAA40430.1; JOINED.
DR   EMBL; M86615; AAA40430.1; JOINED.
DR   EMBL; M86616; AAA40430.1; JOINED.
DR   EMBL; M86617; AAA40430.1; JOINED.
DR   EMBL; M86618; AAA40430.1; JOINED.
DR   EMBL; M86619; AAA40430.1; JOINED.
DR   EMBL; U16175; AAA98537.1; -.
DR   PIR; A46016; A46016.
DR   PIR; A44124; A44124.
```

```
DR   HSSP; P35555; 1EMO.
DR   MGD; MGI:98739; THBS3.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    956       THROMBOSPONDIN 3.
FT   DOMAIN       22    272       N-TERMINAL.
FT   DOMAIN      274    456       4 X EGF-TYPE REPEATS.
FT   DOMAIN      488    719       7 X TSP TYPE-3 REPEATS (CA-BINDING).
FT   DOMAIN      720    956       C-TERMINAL.
FT   DOMAIN      274    315       EGF-LIKE 1.
FT   DOMAIN      316    369       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      370    413       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      414    456       EGF-LIKE 4.
FT   REPEAT      488    523       TSP TYPE-3 1.
FT   REPEAT      524    546       TSP TYPE-3 2.
FT   REPEAT      547    582       TSP TYPE-3 3.
FT   REPEAT      583    605       TSP TYPE-3 4.
FT   REPEAT      606    643       TSP TYPE-3 5.
FT   REPEAT      644    683       TSP TYPE-3 6.
FT   REPEAT      684    719       TSP TYPE-3 7.
FT   DISULFID    266    266       INTERCHAIN (PROBABLE).
FT   DISULFID    269    269       INTERCHAIN (PROBABLE).
FT   DISULFID    278    289       BY SIMILARITY.
FT   DISULFID    283    300       BY SIMILARITY.
FT   DISULFID    303    314       BY SIMILARITY.
FT   DISULFID    320    332       BY SIMILARITY.
FT   DISULFID    326    341       BY SIMILARITY.
FT   DISULFID    344    368       BY SIMILARITY.
FT   DISULFID    374    388       BY SIMILARITY.
FT   DISULFID    382    397       BY SIMILARITY.
FT   DISULFID    400    412       BY SIMILARITY.
FT   DISULFID    418    432       BY SIMILARITY.
FT   DISULFID    426    442       BY SIMILARITY.
FT   DISULFID    444    455       BY SIMILARITY.
FT   CARBOHYD    310    310       POTENTIAL.
FT   CARBOHYD    407    407       POTENTIAL.
FT   CARBOHYD    644    644       POTENTIAL.
FT   CARBOHYD    937    937       POTENTIAL.
FT   CONFLICT    433    433       L -> F (IN REF. 2).
FT   CONFLICT    563    563       N -> D (IN REF. 2).
FT   CONFLICT    720    720       G -> E (IN REF. 2).
FT   CONFLICT    871    871       L -> W (IN REF. 2 AND 4).
SQ   SEQUENCE   956 AA;  103972 MW;  D7645284 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 956;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 RISGVGID 15
          |:||||||
Db    352 RVSGVGID 359


RESULT  13
CBS_RAT
ID   CBS_RAT        STANDARD;      PRT;   560 AA.
AC   P32232;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CYSTATHIONINE BETA-SYNTHASE (EC 4.2.1.22) (SERINE SULFHYDRASE)
DE   (BETA-THIONASE) (HEMOPROTEIN H-450).
GN   CBS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 92283859.
RA   SWAROOP M., BRADLEY K., OHURA T., TAHARA T., ROPER M.D.,
RA   ROSENBERG L.E., KRAUS J.P.;
RT   "Rat cystathionine beta-synthase. Gene organization and alternative
RT   splicing.";
RL   J. Biol. Chem. 267:11455-11461(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-22 AND 38-47.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX   MEDLINE; 91210211.
RA   ISHIHARA S., MOROHASHI K.-I., SADANO H., KAWABATA S.-I., GOTOH O.,
RA   OMURA T.;
RT   "Molecular cloning and sequence analysis of cDNA coding for rat liver
RT   hemoprotein H-450.";
RL   J. Biochem. 108:899-902(1990).
CC   -!- CATALYTIC ACTIVITY: L-SERINE + L-HOMOCYSTEINE = CYSTATHIONINE +
CC       H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: FIRST STEP IN HOMOCYSTEINE TRANSULFURATION.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- ALTERNATIVE PRODUCTS: FOUR VARIANTS CAN BE FORMED BY ALTERNATIVE
CC       SPLICING (TYPE I THROUGH IV). THE SEQUENCE OF TYPE I IS SHOWN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, AND BRAIN.
CC   -!- PTM: BINDS COVALENTLY TO A HEME GROUP THROUGH A THIOLATE
CC       LIGAND.
CC   -!- SIMILARITY: BELONGS TO THE CYSTEINE SYNTHASE/CYSTATHIONINE BETA-
CC       SYNTHASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CBS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88344; AAB02042.1; -.
DR   EMBL; M88346; AAA42024.1; -.
DR   EMBL; D01098; BAA00883.1; -.
DR   PIR; JX0145; JX0145.
DR   PIR; A42790; A42790.
DR   PFAM; PF00291; S_T_dehydratase; 1.
DR   PFAM; PF00571; CBS; 1.
DR   PROSITE; PS00901; CYS_SYNTHASE; 1.
KW   Lyase; Cysteine biosynthesis; Pyridoxal phosphate; Heme; CBS domain;
KW   Alternative splicing.
FT   INIT_MET      0      0
FT   BINDING     115    115       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   DOMAIN      411    464       CBS.
FT   VARSPLIC    513    513       S -> Y (IN TYPE III).
FT   VARSPLIC    514    527       MISSING (IN TYPE III).
FT   VARSPLIC    449    477       AILGMVTLGNMLSSLLAGKVRPSDEVCKV ->
FT                                LRQSKDICHPTKRHIIQAHGLRKVPDTEA (IN TYPE
FT                                IV).
FT   VARSPLIC    478    560       MISSING (IN TYPE III).
FT   CONFLICT    414    414       L -> P (IN REF. 2).
SQ   SEQUENCE   560 AA;  61323 MW;  6D972105 CRC32;


  Query Match             50.0%;  Score 37;  DB 1;  Length 560;
  Best Local Similarity   35.7%;  Pred. No. 37;
  Matches    5;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 TQKITYRISGVGID 15
          |::  | : |:| |
Db    292 TEQTAYEVEGIGYD 305


RESULT  14
DSG2_HUMAN
```

```
ID   DSG2_HUMAN     STANDARD;      PRT;  1117 AA.
AC   Q14126;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 2 PRECURSOR (HDGC).
GN   DSG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CARCINOMA;
RX   MEDLINE; 94192736.
RA   SCHAEFER S., KOCH P.J., FRANKE W.W.;
RT   "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT   expression catalogue of the desmoglein subfamily of desmosomal
RT   cadherins.";
RL   Exp. Cell Res. 211:391-399(1994).
RN   [2]
RP   SEQUENCE OF 777-1117 FROM N.A.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z26317; CAA81226.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 125671; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     48       POTENTIAL.
FT   CHAIN        49   1117       DESMOGLEIN 2.
FT   DOMAIN       49    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    633       POTENTIAL.
FT   DOMAIN      634   1117       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       49    159       CADHERIN 1.
FT   REPEAT      160    272       CADHERIN 2.
FT   REPEAT      273    387       CADHERIN 3.
FT   REPEAT      388    502       CADHERIN 4.
FT   REPEAT      880    911       DESMOGLEIN REPEAT 1.
FT   REPEAT      912    941       DESMOGLEIN REPEAT 2.
FT   REPEAT      942    967       DESMOGLEIN REPEAT 3.
FT   REPEAT      968    991       DESMOGLEIN REPEAT 4.
FT   REPEAT      992   1020       DESMOGLEIN REPEAT 5.
FT   REPEAT     1021   1050       DESMOGLEIN REPEAT 6.
FT   CARBOHYD    111    111       POTENTIAL.
FT   CARBOHYD    181    181       POTENTIAL.
FT   CARBOHYD    308    308       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    513    513       POTENTIAL.
SQ   SEQUENCE   1117 AA;  122385 MW;  84D3B898 CRC32;
```

```
  Query Match          50.0%;  Score 37;  DB 1;  Length 1117;
  Best Local Similarity  63.6%;  Pred. No. 74;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 KITYRISGVGI 14
           |||| : :| ||
Db      83 KITYKYTGKGI 93
```

```
RESULT  15
HMGL_HUMAN
ID   HMGL_HUMAN     STANDARD;      PRT;   325 AA.
AC   P35914;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)
DE   (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLUTARATE-COA LYASE).
GN   HMGCL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93179448.
RA   MITCHELL G.A., ROBERT M.-F., HRUZ P.W., WANG S., FONTAINE G.,
RA   BEHNKE C.E., MENDE-MUELLER L.M., SCHAPPERT K., LEE C., GIBSON K.M.,
RA   MIZIORKO H.M.;
RT   "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL). Cloning of human
RT   and chicken liver HL cDNAs and characterization of a mutation causing
RT   human HL deficiency.";
RL   J. Biol. Chem. 268:4376-4381(1993).
RN   [2]
RP   SEQUENCE OF 21-325 FROM N.A.
RX   MEDLINE; 96207305.
RA   WANG S.P., ROBERT M.-F., GIBSON K.M., WANDERS R.J.A., MITCHELL G.A.;
RT   "3-Hydroxy-3-methylglutaryl CoA lyase (HL): mouse and human HL gene
RT   (HMGCL) cloning and detection of large gene deletions in two
RT   unrelated HL-deficient patients.";
RL   Genomics 33:99-104(1996).
RN   [3]
RP   VARIANT HMG ARG-233.
RX   MEDLINE; 96394619.
RA   ROBERTS J., MITCHELL G.A., MIZIORKO H.M.;
RT   "Modeling of mutatino responsible for human 3-hydroxy-3-
RT   methylglutaryl-CoA lyase deficiency implicates histidine-233 as an
RT   active site residue.";
RL   J. Biol. Chem. 271:24604-24609(1996).
RN   [4]
RP   VARIANTS HMG GLN-41; GLU-42; GLY-42 AND HIS-42.
RX   MEDLINE; 98130531.
RA   MITCHELL G.A., OZAND P.T., ROBERT M.-F., ASHMARINA L., ROBERTS J.,
RA   GIBSON K.M., WANDERS R.J., WANG S., CHEVALIER I., PLOECHL E.,
RA   MIZIORKO H.;
RT   "HMG CoA lyase deficiency: identification of five causal point
RT   mutations in codons 41 and 42, including a frequent Saudi Arabian
RT   mutation, R41Q.";
RL   Am. J. Hum. Genet. 62:295-300(1998).
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA =
CC       ACETYL-COA + ACETOACETATE.
CC   -!- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- TISSUE SPECIFICITY: FIBROBLASTS, LIVER AND LYMPHOBLASTS.
CC   -!- DISEASE: DEFECTS IN HMGCL ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC       DISEASE KNOWN AS HYDROXYMETHYLGLUTARICACIDURIA (HMG) WHICH CAN
CC       LEAD TO HYPOGLYCEMIA AND COMA.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L07033; AAA92733.1; -.
DR   EMBL; U49719; AAB19099.1; -.
DR   EMBL; U49712; AAB19099.1; JOINED.
DR   EMBL; U49713; AAB19099.1; JOINED.
DR   EMBL; U49714; AAB19099.1; JOINED.
DR   EMBL; U49715; AAB19099.1; JOINED.
DR   EMBL; U49716; AAB19099.1; JOINED.
DR   EMBL; U49717; AAB19099.1; JOINED.
DR   EMBL; U49718; AAB19099.1; JOINED.
DR   PIR; A45470; A45470.
DR   MIM; 246450; -.
DR   PFAM; PF00682; HMGL-like; 1.
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase; Mitochondrion; Transit peptide; Disease mutation.
FT   TRANSIT       1     27       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN        28    325       HYDROXYMETHYLGLUTARYL-COA LYASE.
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   VARIANT      41     41       R -> Q (IN HMG).
FT                                /FTId=VAR_003744.
FT   VARIANT      42     42       D -> E (IN HMG).
FT                                /FTId=VAR_003745.
FT   VARIANT      42     42       D -> G (IN HMG).
FT                                /FTId=VAR_003746.
FT   VARIANT      42     42       D -> H (IN HMG).
FT                                /FTId=VAR_003747.
FT   VARIANT      70     70       V -> L (IN HMG).
FT                                /FTId=VAR_003748.
FT   VARIANT     233    233       H -> R (IN HMG).
FT                                /FTId=VAR_003749.
FT   CONFLICT    243    243       T -> A (IN REF. 2).
SQ   SEQUENCE   325 AA;  34390 MW;  DCDBABC8 CRC32;


  Query Match           50.0%;  Score 37;  DB 1;  Length 325;
  Best Local Similarity  42.9%;  Pred. No. 21;
  Matches    6;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ATQKITYRISGVGI 14
           ||: : | : |:||
Db     277 ATEDLVYMLEGLGI 290


Search completed: January  7, 2000, 13:25:35
Job time: 387 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:12:17 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-1
Perfect score: 74
Sequence: 1 ATQKITYRISGVGID 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	87.8	993	11	O35902	O35902 mus musculu
2	40	54.1	2023	5	O96542	O96542 drosophila
3	39	52.7	3380	5	Q24292	Q24292 drosophila
4	39	52.7	441	10	O22282	O22282 arabidopsis
5	38	51.4	147	1	O26829	O26829 methanobact
6	38	51.4	331	2	O66968	O66968 aquifex aeo
7	38	51.4	303	2	P95639	P95639 rhodospiril
8	38	51.4	177	3	Q12036	Q12036 saccharomyc
9	38	51.4	2269	5	O77360	O77360 plasmodium
10	38	51.4	287	5	Q20481	Q20481 caenorhabdi
11	38	51.4	82	7	O77862	O77862 oreochromis
12	38	51.4	1177	10	O23377	O23377 arabidopsis
13	38	51.4	433	10	O81027	O81027 arabidopsis
14	37	50.0	402	2	O83781	O83781 treponema p
15	37	50.0	2408	4	Q92566	Q92566 homo sapien
16	37	50.0	1161	5	Q04901	Q04901 halocynthia
17	37	50.0	63	7	O46802	O46802 poeciliopsi
18	37	50.0	1363	12	Q66290	Q66290 human coron
19	37	50.0	1363	12	Q66291	Q66291 human coron
20	36	48.6	296	2	O31431	O31431 bacillus su
21	36	48.6	482	2	Q59326	Q59326 clostridium
22	36	48.6	507	2	O85888	O85888 sphingomona
23	36	48.6	381	2	Q47594	Q47594 escherichia
24	36	48.6	191	4	O95896	O95896 homo sapien
25	36	48.6	839	5	O62002	O62002 branchiosto
26	36	48.6	3097	5	O15943	O15943 drosophila
27	36	48.6	76	7	O46796	O46796 poeciliopsi
28	36	48.6	82	7	Q31538	Q31538 pseudotroph
29	36	48.6	82	7	Q31536	Q31536 pseudotroph
30	36	48.6	82	7	Q31537	Q31537 pseudotroph
31	36	48.6	1065	10	O48948	O48948 arabidopsis
32	36	48.6	138	12	O90818	O90818 human immun
33	36	48.6	717	13	Q91353	Q91353 gallus gall
34	35	47.3	294	1	O54288	O54288 sulfolobus
35	35	47.3	855	1	O57852	O57852 pyrococcus
36	35	47.3	420	2	O68332	O68332 pseudomonas
37	35	47.3	149	2	Q48688	Q48688 lactococcus
38	35	47.3	780	2	P77846	P77846 clostridium
39	35	47.3	362	2	O69445	O69445 methylococc
40	35	47.3	339	2	O85193	O85193 streptococc
41	35	47.3	531	2	Q9Z9P0	Q9z9p0 bacillus sp
42	35	47.3	878	4	Q15855	Q15855 homo sapien
43	35	47.3	1223	5	O44327	O44327 caenorhabdi
44	35	47.3	432	5	O61820	O61820 caenorhabdi
45	35	47.3	572	5	Q23919	Q23919 dictyosteli

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 87.8%; Score 65; DB 11; Length 993;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QKITYRISGVGID 15
||||||||||||| |
Db 81 QKITYRISGVGID 93

RESULT 2
O96542
ID O96542 PRELIMINARY; PRT; 2023 AA.
AC O96542;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENHANCER OF POLYCOMB.
GN E.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=IMAGINAL DISC;
RX   MEDLINE; 98407961.
RA   STANKUNAS K., BERGER J., RUSE C., SINCLAIR D.A.R., RANDAZZO F.,
RA   BROCK H.W.;
RT   "The Enhancer of Polycomb gene of Drosophila encodes a chromatin
RT   protein conserved in yeast and mammals.";
RL   Development 125:4055-4066(1998).
DR   EMBL; AF079764; AAC64271.1; -.
SQ   SEQUENCE   2023 AA;  220680 MW;  1D7D99F6 CRC32;


  Query Match             54.1%;  Score 40;  DB 5;  Length 2023;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          |:||:|| |   |:
Db   1453 ASQKVTYAIQKAGV 1466


RESULT   3
Q24292
ID   Q24292      PRELIMINARY;      PRT;  3380 AA.
AC   Q24292;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ADHERIN.
GN   DACHSOUS.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95324813.
RA   CLARK H.F., BRENTRUP D., SCHNEITZ K., BIEBER A., GOODMAN C., NOLL M.;
RT   "Dachsous encodes a member of the cadherin superfamily that controls
RT   imaginal disc morphogenesis in Drosophila.";
RL   Genes Dev. 9:1530-1542(1995).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; L08811; AAA79329.1; -.
DR   FLYBASE; FBgn0000497; ds.
DR   PFAM; PF00028; cadherin; 25.
DR   PROSITE; PS00232; CADHERIN; 18.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   3380 AA;  366356 MW;  7D83FAC3 CRC32;


  Query Match             52.7%;  Score 39;  DB 5;  Length 3380;
  Best Local Similarity   58.3%;  Pred. No. 4.4e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 KITYRISGVGID 15
          :|||||   |:|
Db   1017 RITYRIVDAGVD 1028


RESULT   4
O22282
ID   O22282      PRELIMINARY;      PRT;  441 AA.
AC   O22282;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE   T5I7.1 PROTEIN.
GN   T5I7.1 OR F17A14.8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC003000; AAB87120.1; -.
DR   EMBL; AC003674; AAB97126.1; -.
SQ   SEQUENCE   441 AA;  47873 MW;  F462C421 CRC32;


  Query Match             52.7%;  Score 39;  DB 10;  Length 441;
  Best Local Similarity   54.5%;  Pred. No. 44;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 QKITYRISGVG 13
          ||: ||::| |
Db    368 QKLLYRVNGAG 378


RESULT   5
O26829
ID   O26829      PRELIMINARY;      PRT;  147 AA.
AC   O26829;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   HYPOTHETICAL 15.6 KD PROTEIN.
GN   MTH733.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE; 98037514.
RA   SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA   ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA   HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
RA   SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA   JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA   MCDOUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA   DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000852; AAB85238.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   147 AA;  15626 MW;  9029E848 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 147;
  Best Local Similarity   61.5%;  Pred. No. 19;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVG 13
          || |:| |  |||
Db    135 ATVKVTNRTGGVG 147


RESULT   6
O66968
ID   O66968      PRELIMINARY;      PRT;  331 AA.
AC   O66968;
```

```
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE.
GN   PURM.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RA   DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA   GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA   FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE000706; AAC06918.1; -.
DR   PFAM; PF00586; AIRS; 1.
KW   Ligase.
SQ   SEQUENCE   331 AA;  36768 MW;  4C3D13B0 CRC32;

  Query Match             51.4%;  Score 38;  DB 2;  Length 331;
  Best Local Similarity   63.6%;  Pred. No. 48;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 ITYRISGVGID 15
           :||| :|| ||
Db       2 VTYRSAGVDID 12


RESULT   7
P95639
ID   P95639      PRELIMINARY;      PRT;   303 AA.
AC   P95639;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL)
DE   (3-HYDROXY-3-METHYLGLUTARATE-COA LYASE).
GN   HMGL.
OS   Rhodospirillum rubrum.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC   Rhodospirillum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S1;
RX   MEDLINE; 97157062.
RA   BALTSCHEFFSKY M., BROSCHE M., HULTMAN T., LUNDVIK L., NYREN P.,
RA   SAKAI-NORE Y., SEVERIN A., STRID A.;
RT   "A 3-hydroxy-3-methylglutaryl-CoA lyase gene in the photosynthetic
RT   bacterium Rhodospirillum rubrum.";
RL   Biochim. Biophys. Acta 1337:113-122(1997).
CC   -!- CATALYTIC ACTIVITY:
CC       (S)-3-HYDROXY-3-METHYLGLUTARYL-COA = ACETYL-COA + ACETOACETATE.
CC   -!- SIMILARITY: TO OTHER SPECIES HMG-COA LYASE.
DR   EMBL; U41280; AAB50182.1; -.
DR   PFAM; PF00682; HMGL-like; 1.
DR   PROSITE; PS01062; HMG_COA_LYASE; 1.
KW   Lyase.
FT   ACT_SITE    239    239       BY SIMILARITY.
SQ   SEQUENCE   303 AA;  31138 MW;  44C2DA4B CRC32;

  Query Match             51.4%;  Score 38;  DB 2;  Length 303;
  Best Local Similarity   42.9%;  Pred. No. 43;
  Matches    6;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ATQKITYRISGVGI 14
             ||: : | : |:||
Db     250 ATEDVVYMLDGMGI 263


RESULT   8
Q12036
ID   Q12036      PRELIMINARY;      PRT;   177 AA.
AC   Q12036;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE   P2604 PROTEIN (ORF YPL148C).
GN   PPT2.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB972;
RA   PURNELLE B., COMBLEZ S., COSTER F., NAVEAU F., GOFFEAU A.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   STUIBLE H.P., MEIER S., WAGNER C., SCHWEIZER E.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X96770; CAA65545.1; -.
DR   EMBL; Z73504; CAA97853.1; -.
DR   EMBL; Y16253; CAA76138.1; -.
KW   Transferase; Phosphopantetheine.
SQ   SEQUENCE   177 AA;  20398 MW;  2A4460D2 CRC32;

  Query Match             51.4%;  Score 38;  DB 3;  Length 177;
  Best Local Similarity   46.7%;  Pred. No. 23;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATQKITYRISGVGID 15
             |:: |  :|:|||:|
Db       8 ASRNIGRKIAGVGVD 22


RESULT   9
O77360
ID   O77360      PRELIMINARY;      PRT;  2269 AA.
AC   O77360;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MAL3P4.25 PROTEIN.
GN   MAL3P4.25.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   HAMLIN N., LAWSON D., BARRELL B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL008970; CAA15615.1; -.
DR   PFAM; PF00271; helicase_C; 1.
SQ   SEQUENCE   2269 AA;  267233 MW;  11E1C8E0 CRC32;

  Query Match             51.4%;  Score 38;  DB 5;  Length 2269;
  Best Local Similarity   88.9%;  Pred. No. 4.3e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      3 QKITYRISG 11
          ||| |||||
Db    573 QKIGYRISG 581


RESULT  10
Q20481
ID   Q20481      PRELIMINARY;      PRT;   287 AA.
AC   Q20481;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F46G10.3 PROTEIN.
GN   F46G10.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   COLES L.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z50177; CAA90547.1; -.
SQ   SEQUENCE   287 AA;  32464 MW;  F4C7797C CRC32;


  Query Match             51.4%;  Score 38;  DB 5;  Length 287;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGV 12
          || |: |||| |
Db    272 ATMKLDYRISDV 283


RESULT  11
O77862
ID   O77862      PRELIMINARY;      PRT;    82 AA.
AC   O77862;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MHC CLASS II B LOCUS 4 (FRAGMENT).
OS   Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC   Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC   Perciformes; Labroidei; Cichlidae; Tilapia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98315113.
RA   MALAGA-TRILLO E., MCANDREW B., VINCEK V., ZALESKA-RUTCZYNSKA Z.,
RA   SUELTMANN H., FIGUEROA F., KLEIN J.;
RT   "Linkage relationships and haplotype polymorphism among cichlid Mhc
RT   class II B loci.";
RL   Genetics 149:1527-1547(1998).
DR   EMBL; AF049971; AAC41310.1; -.
DR   PFAM; PF00969; MHC_II_beta; 1.
KW   MHC.
FT   NON_TER       1      1
FT   NON_TER      82     82
SQ   SEQUENCE   82 AA;  9607 MW;  D1718622 CRC32;


  Query Match             51.4%;  Score 38;  DB 7;  Length 82;
  Best Local Similarity   53.8%;  Pred. No. 9.8;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 QKITYRISGVGID 15
          || || :: :|||
Db     70 QKETYCLNNIGID 82


RESULT  12
O23377
ID   O23377      PRELIMINARY;      PRT;  1177 AA.
AC   O23377;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PDR5-LIKE ABC TRANSPORTER.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BEVAN M., STIEKEMA W., MURPHY G., WAMBUTT R., POHL T., TERRYN N.,
RA   KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,
RA   PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,
RA   JONES J., PALME K., ANSORGE W., DELSENY M., BANCROFT I., MEWES H.W.,
RA   SCHUELLER C., CHALWATZIS N.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97338; CAB10302.1; -.
DR   PFAM; PF00005; ABC_tran; 1.
SQ   SEQUENCE   1177 AA;  134263 MW;  32946ADC CRC32;


  Query Match             51.4%;  Score 38;  DB 10;  Length 1177;
  Best Local Similarity   69.2%;  Pred. No. 2e+02;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 QKITYRISGVGID 15
          |||  ||  ||||
Db   1042 QKIRKRIDEVGID 1054


RESULT  13
O81027
ID   O81027      PRELIMINARY;      PRT;   433 AA.
AC   O81027;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE HYDROXYMETHYLGLUTARYL-COA LYASE PRECURSOR.
GN   F12C20.16.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., RONNING C.M., LIN X., KETCHUM K.A., CROSBY M.L.,
RA   BRANDON R.C., SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R.,
```

```
RA   ADAMS M.D., SOMERVILLE C.R., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005168; AAC32247.1; -.
DR   PFAM; PF00682; HMGL-like; 1.
KW   Lyase.
SQ   SEQUENCE   433 AA;  46385 MW;  A449D9B9 CRC32;


  Query Match             51.4%;  Score 38;  DB 10;  Length 433;
  Best Local Similarity   35.7%;  Pred. No. 65;
  Matches    5;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||: : | ::|:|:
Db    377 ATEDVVYMLNGLGV 390


RESULT  14
O83781
ID   O83781      PRELIMINARY;      PRT;   402 AA.
AC   O83781;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 44.6 KD PROTEIN.
GN   TP0803.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98332770.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001251; AAC65775.1; -.
DR   PFAM; PF01368; DHH; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   402 AA;  44577 MW;  C237C707 CRC32;


  Query Match             50.0%;  Score 37;  DB 2;  Length 402;
  Best Local Similarity   57.1%;  Pred. No. 90;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGI 14
          ||  |   |||:||
Db    309 ATSNIAESISGIGI 322


RESULT  15
Q92566
ID   Q92566      PRELIMINARY;      PRT;  2408 AA.
AC   Q92566;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   MYELOBLAST KIAA0279 (FRAGMENT).
GN   KIAA0279.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97191544.
RA   NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,
RA   TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; D87469; BAA13407.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF00008; EGF; 6.
DR   PFAM; PF00054; laminin_G; 1.
DR   PROSITE; PS00232; CADHERIN; 4.
KW   Glycoprotein; Cell adhesion; Transmembrane; Calcium-binding; Repeat.
FT   NON_TER       1      1
SQ   SEQUENCE   2408 AA;  261740 MW;  CDBA2001 CRC32;


  Query Match             50.0%;  Score 37;  DB 4;  Length 2408;
  Best Local Similarity   50.0%;  Pred. No. 6.9e+02;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 KITYRISGVGID 15
          :: ||::||| |
Db     24 RLEYRLAGVGHD 35


Search completed: January 12, 2000, 23:15:18
Job time: 181 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:34 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-2
Perfect score: 78
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	100.0	999	1	R30742	Human pemphigus vu
2	78	100.0	30	1	R93960	Peptide combining
3	78	100.0	614	1	W07908	Pemphigus vulgaris
4	78	100.0	15	1	W04842	Self epitope of de
5	64	82.1	19	1	R93961	Peptide combining
6	64	82.1	21	1	R93962	Peptide combining
7	56	71.8	560	1	W13009	Segment of desmoso
8	55	70.5	778	1	W15489	Pemphigus foliaceu
9	48	61.5	17	1	R93963	Peptide combining
10	42	53.8	713	1	W13136	Full length human
11	42	53.8	713	1	W25638	Human cadherin-13.
12	41	52.6	794	1	W13135	Putative human cad
13	41	52.6	794	1	W25637	Human cadherin-12.
14	39	50.0	1091	1	R28822	Alpha 6B integrin
15	39	50.0	1073	1	R55272	Alpha subunit of i
16	38	48.7	797	1	R86866	Human protocadheri
17	38	48.7	433	1	W41732	Arabidopsis chloro
18	38	48.7	341	1	W53939	L.lactis MG1363 or
19	37	47.4	863	1	R22357	Carrot aspartokina
20	37	47.4	878	1	R55060	Sequence of human
21	37	47.4	499	1	R60609	Tobamovirus replic
22	37	47.4	845	1	R70065	Hepatitis B virus
23	37	47.4	878	1	R85487	Human E-cadherin p
24	37	47.4	616	1	R86867	Rat protocadherin
25	37	47.4	982	1	W20980	H. pylori secreted
26	37	47.4	863	1	W75440	Carrot aspartokina
27	37	47.4	978	1	Y10992	H. pylori ORF 06ep
28	36	46.2	566	1	R13228	Endoglucanase enco
29	36	46.2	1203	1	R58911	Product of alterna
30	36	46.2	1026	1	R58906	Human protocadheri
31	36	46.2	515	1	R69505	Aspergillus sp. re
32	36	46.2	511	1	R69506	Aspergillus sp. re
33	36	46.2	515	1	R69504	Aspergillus sp. re
34	36	46.2	883	1	R69627	Bacteriophage T7 R
35	36	46.2	1203	1	R87152	Alternatively spli
36	36	46.2	1026	1	R87146	Protocadherin clon
37	36	46.2	488	1	R85875	WD-40 domain-contg
38	36	46.2	31	1	R85793	Peptide rI from th
39	36	46.2	954	1	W19752	Yeast inhibitor of
40	36	46.2	20	1	W40250	Human wild-type E-
41	36	46.2	52	1	Y11246	Streptococcus pneu
42	36	46.2	883	1	Y01322	Wild-type T7 RNA p
43	36	46.2	883	1	Y01353	Wild-type T7 RNA p
44	36	46.2	883	1	Y01286	Wild-type RNA poly
45	35	44.9	712	1	R27823	Sequence encoded b

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 78; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDINI 15
||||||||||||||||
Db 97 FGIFVVDKNTGDINI 111

RESULT 2
R93960
ID R93960 standard; peptide; 30 AA.
AC R93960;
DT 05-JUL-1996 (first entry)
DE Peptide combining with anti-interepidermal cellular antibody.
KW anti-interepidermal cellular antibody; autoantibody; adsorbent.
OS Synthetic.
PN J07309893-A.
PD 28-NOV-1995.
PF 18-MAY-1994; 129556.
PR 18-MAY-1994; JP-129556.
PA (KURS) KURARAY CO LTD.
DR WPI; 96-045392/05.
PT Anti-interepidermal cellular antibody-combining peptide - which can
PT be immobilised on column to form adsorbent useful for treating
PT diseases related to the antibody
PS Claim 1; Page 2; 7pp; Japanese.
CC A new peptide is disclosed which contains at least 5 contiguous amino
CC acids from the sequence Val Gly Ile Asp Gln Pro Pro Phe Gly Ile Phe
CC Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC Glu Glu (the present sequence), the peptide not containing more than
CC 50 residues. The peptide combines with anti-interepidermal cellular
CC antibody. It can be immobilised on a carrier to prepare an absorbent
CC useful for the treatment of diseases related to anti-interepidermal
CC cellular antibody.

```
SQ   Sequence   30 AA;


  Query Match             100.0%;  Score 78;  DB 1;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||||||||||||
Db      8 FGIFVVDKNTGDINI 22


RESULT   3
W07908
ID   W07908 standard; protein; 614 AA.
AC   W07908;
DT   29-JAN-1997  (first entry)
DE   Pemphigus vulgaris antigen protein extracellular region.
KW   Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW   treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW   dermatology.
OS   Homo sapiens.
PN   J08188540-A.
PD   23-JUL-1996.
PF   30-JUN-1995; 165632.
PR   30-JUN-1994; JP-173291.
PA   (NISH/) NISHUKAWA T.
DR   WPI; 96-388562/39.
PT   Fused protein recognised by pemphigus vulgaris auto:antibody -
PT   useful to treat and diagnose pemphis vulgaris
PS   Claim 1; Page 7-9; 9pp; Japanese.
CC   W07908 represents the human pemphigus vulgaris (PV) antigen
CC   extracellular region. The PV antigen is produced in patients with
CC   pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC   relapsing disease causing suprabasal, intra-epidermal bullae
CC   (vesicles) of the skin and mucuous membranes, which is fatal if
CC   untreated. The PV antigen was fused to a human IgG1 hinge region
CC   and the resulting fusion protein is useful to treat or diagnose
CC   pemphigus vulgaris.
SQ   Sequence   614 AA;


  Query Match             100.0%;  Score 78;  DB 1;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||||||||||||
Db     96 FGIFVVDKNTGDINI 110


RESULT   4
W04842
ID   W04842 standard; peptide; 15 AA.
AC   W04842;
DT   18-FEB-1997  (first entry)
DE   Self epitope of desmoglein 3, implicated in autoimmune disease.
KW   Tolerisation; self-epitope; antigen; autoimmune disease;
KW   autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW   pemphigus vulgaris; desmoglein; multiple sclerosis;
KW   herpes simplex virus; adenovirus; phosphomannomutase;
KW   human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW   influenza; haemagglutinin; reovirus; sigma protein.
OS   Homo sapiens.
PN   WO9627387-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03182.
PR   07-MAR-1995; US-400796.
PA   (HARD ) HARVARD COLLEGE.
PI   Strominger JL,  Wucherpfennig KW;
DR   WPI; 96-425218/42.
PT   Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT   antigens - useful in disease treatment, and method for
PT   identification of other self and non-self antigens implicated in
PT   auto-immune disease
PS   Claim 1; Page 38; 58pp; English.
CC   Pharmaceutical preparations for tolerisation to antigens comprise
CC   either an isolated human non-collagen or non-myslin basic protein
CC   (MBP) polypeptide which is capable of tolerising an individual to an
CC   autoantigen; or an isolated human pathogen polypeptide capable of
CC   tolerising an individual to that polypeptide.  In both cases, the
CC   polypeptide (whether self or non-self) includes an amino acid
CC   sequence corresponding to a sequence motif for a MHC class II
CC   protein, such as HLA-DR, which is associated with a human autoimmune
CC   disease and which binds to the polypeptide to activate autoreactive
CC   T-cells in individuals with the autoimmune disease. This peptide is
CC   derived from the human desmoglein 3 protein (amino acids 97-111)
CC   and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC   derived from the human desmoglein protein are described in W04841-47.
SQ   Sequence   15 AA;


  Query Match             100.0%;  Score 78;  DB 1;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-08;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||||||||||||
Db      1 FGIFVVDKNTGDINI 15


RESULT   5
R93961
ID   R93961 standard; peptide; 19 AA.
AC   R93961;
DT   05-JUL-1996  (first entry)
DE   Peptide combining with anti-interepidermal cellular antibody.
KW   anti-interepidermal cellular antibody; autoantibody; adsorbent.
OS   Synthetic.
PN   J07309893-A.
PD   28-NOV-1995.
PF   18-MAY-1994; 129556.
PR   18-MAY-1994; JP-129556.
PA   (KURS ) KURARAY CO LTD.
DR   WPI; 96-045392/05.
PT   Anti-interepidermal cellular antibody-combining peptide - which can
PT   be immobilised on column to form adsorbent useful for treating
PT   diseases related to the antibody
PS   Example 1; Page 4; 7pp; Japanese.
CC   New peptides are disclosed which contain at least 5 contiguous amino
CC   acids from the sequence Val Gly Ile Asp Gln Pro Pro Phe Gly Ile Phe
CC   Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC   Glu Glu (see R93960), the peptide not containing more than 50 residues.
CC   The peptide combines with anti-interepidermal cellular antibody. It can
CC   be immobilised on a carrier to prepare an absorbent useful for the
CC   treatment of diseases related to anti-interepidermal cellular antibody.
CC   The present sequence is a specific example of the new peptides.
SQ   Sequence   19 AA;


  Query Match              82.1%;  Score 64;  DB 1;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGD 12
          ||||||||||||
Db      8 FGIFVVDKNTGD 19


RESULT   6
R93962
ID   R93962 standard; peptide; 21 AA.
AC   R93962;
DT   05-JUL-1996  (first entry)
DE   Peptide combining with anti-interepidermal cellular antibody.
KW   anti-interepidermal cellular antibody; autoantibody; adsorbent.
```

```
OS   Synthetic.
PN   J07309893-A.
PD   28-NOV-1995.
PF   18-MAY-1994; 129556.
PR   18-MAY-1994; JP-129556.
PA   (KURS ) KURARAY CO LTD.
DR   WPI; 96-045392/05.
PT   Anti-interepidermal cellular antibody-combining peptide - which can
PT   be immobilised on column to form adsorbent useful for treating
PT   diseases related to the antibody
PS   Example 2; Page 5; 7pp; Japanese.
CC   New peptides are disclosed which contain at least 5 contiguous amino
CC   acids from the sequence Val Gly Ile Asp Gln Pro Pro Phe Gly Ile Phe
CC   Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC   Glu Glu (see R93960), the peptide not containing more than 50 residues.
CC   The peptide combines with anti-interepidermal cellular antibody. It can
CC   be immobilised on a carrier to prepare an absorbent useful for the
CC   treatment of diseases related to anti-interepidermal cellular antibody.
CC   The present sequence is a specific example of the new peptides.
SQ   Sequence   21 AA;


  Query Match             82.1%;  Score 64;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGD 12
          ||||||||||||
Db     10 FGIFVVDKNTGD 21


RESULT   7
W13009
ID   W13009 standard; protein; 560 AA.
AC   W13009;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 7; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
CC   carcinomas.
SQ   Sequence   560 AA;


  Query Match             71.8%;  Score 56;  DB 1;  Length 560;
  Best Local Similarity   60.0%;  Pred. No. 0.039;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          |||||  :|:||::|:
Db     50 FGIFVFNKDTGELNV 64


RESULT   8
W15489
ID   W15489 standard; Protein; 778 AA.
AC   W15489;
DT   17-JUN-1997  (first entry)
DE   Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW   Pemphigus foliaceus; autoantibody; constant region; IgG;
KW   extracellular region; antigen; hinge portion; skin;
KW   dermatitis herpetiformis; fusion protein; detection; ss.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1. .545
FT                   /note= "Pemphigus foliaceus antigen protein"
PN   J09077800-A.
PD   25-MAR-1997.
PF   12-SEP-1995; 260899.
PR   12-SEP-1995; JP-260899.
PA   (NISH/) NISHIKAWA T.
DR   WPI; 97-241758/22.
DR   P-PSDB; T66428.
PT   Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT   through the hinge region used to treat pemphigus foliaceus
PS   Claim 1; Page 10-12; 17pp; Japanese.
CC   This sequence represents a fused protein recognised by pemphigus
CC   foliaceus patient autoantibody which comprises the constant region
CC   of IgG linked to the extracellular region of pemphigus foliaceus
CC   antigen protein through the hinge portion.  Pemphigus foliaceus is
CC   a chronic, generalised, vesicular and scaling skin eruption similar
CC   to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC   protein is useful to treat pemphigus foliaceus.  The antigen is
CC   especially administered through an adsorbent upon which the fusion
CC   protein is immobilised via a carrier. The fusion protein is also
CC   useful for detecting pemphigus foliaceus antibodies which is useful
CC   in immunodiagnosis.  The fusion protein has little or no side effects.
SQ   Sequence   778 AA;


  Query Match             70.5%;  Score 55;  DB 1;  Length 778;
  Best Local Similarity   60.0%;  Pred. No. 0.085;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 FGIFVVDKNTGDINI 15
          :|||||:: ||:|||
Db     97 YGIFVINQKTGEINI 111


RESULT   9
R93963
ID   R93963 standard; peptide; 17 AA.
AC   R93963;
DT   05-JUL-1996  (first entry)
DE   Peptide combining with anti-interepidermal cellular antibody.
KW   anti-interepidermal cellular antibody; autoantibody; adsorbent.
OS   Synthetic.
PN   J07309893-A.
PD   28-NOV-1995.
PF   18-MAY-1994; 129556.
PR   18-MAY-1994; JP-129556.
PA   (KURS ) KURARAY CO LTD.
DR   WPI; 96-045392/05.
PT   Anti-interepidermal cellular antibody-combining peptide - which can
PT   be immobilised on column to form adsorbent useful for treating
PT   diseases related to the antibody
PS   Example 3; Page 5; 7pp; Japanese.
CC   New peptides are disclosed which contain at least 5 contiguous amino
CC   acids from the sequence Val Gly Ile Asp Gln Pro Pro Phe Gly Ile Phe
CC   Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC   Glu Glu (see R93960), the peptide not containing more than 50 residues.
CC   The peptide combines with anti-interepidermal cellular antibody. It can
```

```
CC   be immobilised on a carrier to prepare an absorbent useful for the
CC   treatment of diseases related to anti-interepidermal cellular antibody.
CC   The present sequence is a specific example of the new peptides.
SQ   Sequence   17 AA;


  Query Match             61.5%;  Score 48;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.019;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 DKNTGDINI 15
          |||||||||
Db      1 DKNTGDINI 9


RESULT  10
W13136
ID   W13136 standard; Protein; 713 AA.
AC   W13136;
DT   14-MAY-1997  (first entry)
DE   Full length human cadherin-13.
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion.
OS   Homo sapiens.
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   N-PSDB; T61927.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Example 2; Columns 107-112; 59pp; English.
CC   The present sequence is full length human cadherin-13, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a foetal brain cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   713 AA;


  Query Match             53.8%;  Score 42;  DB 1;  Length 713;
  Best Local Similarity   42.9%;  Pred. No. 15;
  Matches    6;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 GIFVVDKNTGDINI 15
          ||| :::||| :::
Db    186 GIFRINENTGSVSV 199


RESULT  11
W25638
ID   W25638 standard; Protein; 713 AA.
AC   W25638;
DT   03-NOV-1997  (first entry)
DE   Human cadherin-13.
KW   Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW   superfamily; cytoskeleton; eatenin; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 703
FT                   /note= "Encoded by CTG"
PN   US5646250-A.
PD   08-JUL-1997.
PF   17-APR-1992; 872643.
PR   19-APR-1993; US-049460.
PR   17-APR-1992; US-872643.
PR   01-NOV-1994; US-332638.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-362997/33.
DR   N-PSDB; T85405.
PT   Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PS   Claim 1; Column 99-102; 56pp; English.
CC   This sequence represents human cadherin-13.  The invention specifically
CC   provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC   cadherin-8, -5, -11 and -13.  Cadherins are calcium-dependent cell
CC   adhesion proteins.  They are glycosylated integral membrane proteins
CC   that have an N-terminal extracellular domain that determines binding
CC   specificity, a hydrophobic membrane spanning region and a C-terminal
CC   cytoplasmic domain, which is highly conserved among members of the
CC   superfamily.  The C-terminal domain interacts with the cytoskeleton
CC   through eatenins and other cytoskeleton-associated proteins.  The
CC   novel cadherin proteins may be used in the analysis of the role of
CC   cadherins in various cancers.  Sequence analysis of the cadherin
CC   proteins also allows investigation of the structure and function of
CC   cadherin.  The cadherin proteins may be isolated by using anti-cadherin
CC   antibodies.  These antibodies may also be used to modulate the activity
CC   of cadherin and to determine the tissue specific distribution of cadherin
CC   proteins.  Each subclass of cadherins has a unique tissue distribution
CC   pattern.
SQ   Sequence   713 AA;


  Query Match             53.8%;  Score 42;  DB 1;  Length 713;
  Best Local Similarity   42.9%;  Pred. No. 15;
  Matches    6;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 GIFVVDKNTGDINI 15
          ||| :::||| :::
Db    186 GIFRINENTGSVSV 199


RESULT  12
W13135
ID   W13135 standard; Protein; 794 AA.
AC   W13135;
DT   14-MAY-1997  (first entry)
DE   Putative human cadherin-12.
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion; putative.
OS   Homo sapiens.
PN   US5597725-A.
PD   28-JAN-1997.
PF   17-APR-1992; 872643.
PR   17-APR-1992; US-872643.
PR   19-APR-1993; US-049460.
PR   26-JAN-1994; US-188228.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-108328/10.
DR   N-PSDB; T61926.
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
PS   Example 2; Columns 101-106; 59pp; English.
CC   The present sequence is a putative human cadherin-12, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a foetal brain cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
SQ   Sequence   794 AA;


  Query Match             52.6%;  Score 41;  DB 1;  Length 794;
  Best Local Similarity   50.0%;  Pred. No. 26;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy      3 IFVVDKNTGDIN 14
          :| :|: ||||:
Db    101 VFTIDETTGDIH 112


RESULT  13
W25637
ID   W25637 standard; Protein; 794 AA.
AC   W25637;
DT   03-NOV-1997  (first entry)
DE   Human cadherin-12.
KW   Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW   superfamily; cytoskeleton; eatenin; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 287
FT                   /note= "Encoded by GCT"
FT   Misc_difference 778
FT                   /note= "Encoded by TTG"
PN   US5646250-A.
PD   08-JUL-1997.
PF   17-APR-1992; 872643.
PR   19-APR-1993; US-049460.
PR   17-APR-1992; US-872643.
PR   01-NOV-1994; US-332638.
PA   (DOHE-) DOHENY EYE INST.
PI   Suzuki S;
DR   WPI; 97-362997/33.
DR   N-PSDB; T85404.
PT   Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PS   Claim 1; Column 95-100; 56pp; English.
CC   This sequence represents human cadherin-12.  The invention specifically
CC   provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC   cadherin-8, -5, -11 and -13.  Cadherins are calcium-dependent cell
CC   adhesion proteins.  They are glycosylated integral membrane proteins
CC   that have an N-terminal extracellular domain that determines binding
CC   specificity, a hydrophobic membrane spanning region and a C-terminal
CC   cytoplasmic domain, which is highly conserved among members of the
CC   superfamily.  The C-terminal domain interacts with the cytoskeleton
CC   through eatenins and other cytoskeleton-associated proteins.  The
CC   novel cadherin proteins may be used in the analysis of the role of
CC   cadherins in various cancers.  Sequence analysis of the cadherin
CC   proteins also allows investigation of the structure and function of
CC   cadherin.  The cadherin proteins may be isolated by using anti-cadherin
CC   antibodies.  These antibodies may also be used to modulate the activity
CC   of cadherin and to determine the tissue specific distribution of cadherin
CC   proteins.  Each subclass of cadherins has a unique tissue distribution
CC   pattern.
SQ   Sequence   794 AA;

  Query Match           52.6%;  Score 41;  DB 1;  Length 794;
  Best Local Similarity   50.0%;  Pred. No. 26;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 IFVVDKNTGDIN 14
          :| :|: ||||:
Db    101 VFTIDETTGDIH 112


RESULT  14
R28822
ID   R28822 standard; Protein; 1091 AA.
AC   R28822;
DT   23-MAR-1993  (first entry)
DE   Alpha 6B integrin subunit.
KW   Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;
KW   extracellular matrix; cytoskeleton; heterodimer; laminin receptor;
KW   immunoprecipitation; JAR; choriocarcinoma.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1012. .1037
FT                   /label= Transmembrane_domain
FT   peptide         1. .23
FT                   /label= Signal_peptide
FT   protein         24. .1091
FT                   /label= Mature_protein
FT   modified_site   223
FT                   /label= Glycosylation_site
FT   modified_site   284
FT                   /label= Glycosylation_site
FT   modified_site   370
FT                   /label= Glycosylation_site
FT   modified_site   513
FT                   /label= Glycosylation_site
FT   modified_site   731
FT                   /label= Glycosylation_site
FT   modified_site   748
FT                   /label= Glycosylation_site
FT   modified_site   891
FT                   /label= Glycosylation_site
FT   modified_site   927
FT                   /label= Glycosylation_site
FT   modified_site   958
FT                   /label= Glycosylation_site
FT   binding_site    230. .238
FT                   /note= "Putative cation binding domain"
FT   binding_site    324. .332
FT                   /note= "Putative cation binding domain"
FT   binding_site    386. .394
FT                   /note= "Putative cation binding domain"
FT   binding_site    441. .449
FT                   /note= "Putative cation binding domain"
FT   domain          1040. .1044
FT                   /label= Cytoplasmic_domain
FT                   /note= "Conservered in virtually all integrins"
FT   misc_difference 1044. .1045
FT                   /note= "Position of deletion cf alpha 6A"
PN   WO9219647-A.
PD   12-NOV-1992.
PF   27-APR-1992; U03527.
PR   03-MAY-1991; US-695564.
PA   (SCRI ) SCRIPPS RES INST.
PI   Quaranta V, Tamura RN;
DR   WPI; 92-398799/48.
DR   N-PSDB; Q31189.
PT   Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
PT   for prodn. of antibodies and in detection of integrin sub-units
PT   in body samples
PS   Disclosure; Page 78-82; 115pp; English.
CC   The sequences given in R28821-22 are the human alpha 6A and 6B
CC   integrin subunits.  Integrins are a family of cell surface receptors
CC   which serve cellular adhesion functions.  These receptors form a link
CC   between the extracellular matrix and the cytoskeleton through their
CC   binding to various extracellular components.  Each integrin receptor
CC   is a heterodimer comprised of an alpha and a beta subunit.  Each alpha
CC   subunit tends to associate with only one type of beta subunit but
CC   there are several exceptions to this rule.  These integrins correspond
CC   to the laminin receptor. The cytoplasmic domain of the 6A and 6B
CC   integrins differs from previously isolated alpha 6 integrins.  The
CC   human alpha 6B was isolated from human choriocarcinoma cell line JAR
CC   by immunoprecipitation studies.
SQ   Sequence   1091 AA;

  Query Match           50.0%;  Score 39;  DB 1;  Length 1091;
  Best Local Similarity   71.4%;  Pred. No. 83;
  Matches   10;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    1;

Qy      1 FGIFVVDKNTGDIN 14
          ||| |  || ||||
Db    377 FGIAV--KNIGDIN 388


RESULT  15
```

```
R55272
ID   R55272 standard; Protein; 1073 AA.
AC   R55272;
DT   27-JAN-1995  (first entry)
DE   Alpha subunit of integrin cell surface receptor.
KW   Integrin; alpha; beta; subunit; glycoprotein; heterodimer;
KW   transmembrane; extracellular matrix; cell signalling; cytoskeleton;
KW   behaviour; signal transduction; receptor.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_binding    230. .238
FT                   /note= "Sequence weakly homologous with cation
FT                   binding site of other integrin alpha
FT                   subunits."
FT   misc_binding    324. .332
FT                   /note= "Putative cation binding site."
FT   misc_binding    386. .394
FT                   /note= "Putative cation binding site."
FT   misc_binding    441. .449
FT                   /note= "Putative cation binding site."
FT   misc_feature    899. .903
FT                   /note= "This sequence closely resembles the
FT                   cleavage site of other integrin O chains."
FT   misc_feature    1040. .1044
FT                   /*tag= f
FT                   /note= "This sequence is conserved in all but the
FT                   Drosophila PS2 alpha subunit"
PN   US5320942-A.
PD   14-JUN-1994.
PF   19-FEB-1987; 016552.
PR   19-FEB-1987; US-016552.
PR   04-JAN-1989; US-293384.
PR   01-OCT-1990; US-591105.
PA   (KAJI/) KAJIJI S.
PA   (QUAR/) QUARANTA V.
PI   Kajiji S,  Quaranta V;
DR   WPI; 94-191533/23.
DR   N-PSDB; Q65673.
PT   Diagnosing presence of abnormal epithelial tissue in vitro -
PT   utilises monoclonal antibodies to alpha6 beta4 cell surface protein
PS   Example 5; Figure 6; 34pp; English.
CC   Integrins are heterodimers comrised of alpha and beta subunits, that
CC   are non-covalently associated transmembrane glycoproteins.  11 alpha
CC   chains and 6 beta chains have been recognised in man.  Each alpha
CC   subunit tends to associate with only one type of beta subunit but
CC   there are several exceptions.  Integrins mediate (in part) the
CC   interaction of cells with the extracellular matrix, forming a link
CC   between the extracellular matrix and the cytoskeleton.  They may
CC   transmit signals from the extracellular to the intracellular
CC   environment, affecting cell behaviour.  This sequence is the alpha6
CC   subunit of an alpha6 beta4 integrin.
SQ   Sequence   1073 AA;


  Query Match             50.0%;  Score 39;  DB 1;  Length 1073;
  Best Local Similarity   71.4%;  Pred. No. 81;
  Matches   10;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 FGIFVVDKNTGDIN 14
           ||| |  || ||||
Db     377 FGIAV--KNIGDIN 388


Search completed: January  6, 2000, 15:40:34
Job time: 1681 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:37 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-2
Perfect score: 78
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	100.0	999	1	IJHUG3	desmoglein 3 precu
2	56	71.8	1117	2	S38673	desmoglein 2 - hum
3	55	70.5	1049	1	IJHUG1	desmoglein 1 precu
4	55	70.5	1043	1	IJBOG1	desmoglein 1 precu
5	51	65.4	785	2	I50180	cadherin-7 - chick
6	45	57.7	761	1	IJBODE	desmocollin 1a - b
7	45	57.7	839	1	IJBODF	desmocollin 1b pre
8	45	57.7	770	2	B48910	desmocollin 1b pre
9	45	57.7	824	2	A48910	desmocollin 1a pre
10	45	57.7	840	2	I37281	Dscla precursor -
11	45	57.7	894	2	I37282	Dsc1b precursor -
12	43	55.1	790	2	I37016	cadherin-6 - human
13	43	55.1	789	2	I52701	K-cadherin - rat
14	43	55.1	790	2	I50178	cadherin-6B - chic
15	43	55.1	790	2	G02678	cadherin-14 - huma
16	43	55.1	214	2	S76379	hypothetical prote
17	43	55.1	566	2	S64005	hypothetical prote
18	42	53.8	713	2	B38992	cadherin 13 precur
19	42	53.8	1072	2	T00041	BH-protocadherin P
20	42	53.8	1200	2	T00042	BH-protocadherin P
21	42	53.8	1069	2	T00040	BH-protocadherin P
22	41	52.6	794	2	I59372	cadherin 12 - huma
23	40	51.3	266	2	A46610	multicatalytic end
24	40	51.3	1626	2	A39242	DNA topoisomerase
25	40	51.3	1612	2	S59969	DNA topoisomerase
26	40	51.3	344	2	H72213	N utilization subs
27	40	51.3	598	2	D69292	aldehyde ferredoxi
28	40	51.3	1715	2	JE0128	Bombyx mori recept
29	39	50.0	790	2	I51638	F-cadherin - Afric
30	39	50.0	165	2	B64121	nonheme ferritin h
31	39	50.0	646	2	S36586	E1 protein - human
32	39	50.0	1358	2	S33653	probable serine/th
33	39	50.0	1073	2	B36429	integrin alpha-6 c
34	39	50.0	1091	2	A41543	integrin alpha-6 c
35	39	50.0	73	2	I51528	integrin alpha 6 s
36	39	50.0	1748	2	S63127	probable membrane
37	38.5	49.4	798	2	S62791	probable lipoprote
38	38	48.7	881	1	JDVLS	DNA-directed DNA p
39	38	48.7	887	1	IJCHCL	E-cadherin precurs
40	38	48.7	657	1	W1WL18	E1 protein - human
41	38	48.7	456	2	S04079	nitrogenase (EC 1.
42	38	48.7	273	2	B64082	outer membrane pro
43	38	48.7	423	2	T06774	cell division prot
44	38	48.7	828	2	S74450	ferrichrome-iron r
45	38	48.7	626	2	C70104	DNA topoisomerase

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 FGIFVVDKNTGDINI 15
           |||||||||||||||
Db      97 FGIFVVDKNTGDINI 111
```

RESULT 2
S38673
desmoglein 2 - human
N;Alternate names: desmoglein HDGC
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C;Accession: S38673; B38872
R;Zimbelmann, R.
submitted to the EMBL Data Library, September 1993
A;Reference number: S38673
A;Accession: S38673
A;Status: preliminary

A;Molecule type: mRNA
A;Residues: 1-1117 <ZIM>
A;Cross-references: EMBL:Z26317; NID:g416177; PIDN:CAA81226.1; PID:g416178
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide
A;Reference number: A38872; MUID:92037656
A;Accession: B38872
A;Molecule type: mRNA
A;Residues: 777-1117 <KOC>
A;Cross-references: GB:S64273
C;Genetics:
A;Gene: GDB:DSG2
A;Cross-references: GDB:128808; OMIM:125671
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein
F;51-158/Domain: cadherin repeat homology <CR1>
F;161-271/Domain: cadherin repeat homology <CR2>

Query Match 71.8%; Score 56; DB 2; Length 1117;
Best Local Similarity 60.0%; Pred. No. 0.19;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDINI 15
|||||:|:||::|:
Db 98 FGIFVFNKDTGELNV 112

RESULT 3
IJHUG1
desmoglein 1 precursor - human
N;Alternate names: desmosomal glycoprotein I
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S16906; A39706; A61254; A61279; S16158
R;Buxton, R.S.
submitted to the EMBL Data Library, November 1990
A;Reference number: S16906
A;Accession: S16906
A;Molecule type: mRNA
A;Residues: 1-1049 <BUX>
A;Cross-references: EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
R;Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; R
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991
A;Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,
A;Reference number: A39706; MUID:91271279
A;Accession: A39706
A;Molecule type: mRNA
A;Residues: 24-1049 <WHE>
A;Cross-references: GB:X56654
R;Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
J. Cell Sci. 99, 809-821, 1991
A;Title: Structural analysis and expression of human desmoglein: a cadherin-like compone
A;Reference number: A61254; MUID:92121251
A;Accession: A61254
A;Molecule type: mRNA
A;Residues: 26-1049 <NIL>
R;Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee
Biochem. Soc. Trans. 19, 1060-1064, 1991
A;Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfami
A;Reference number: A61279; MUID:92175187
A;Accession: A61279
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-55 <WH3>
C;Genetics:
A;Gene: GDB:DSG1
A;Cross-references: GDB:126563; OMIM:125670
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1049/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-493/Domain: cadherin repeat homology <CR4>
F;509-530/Region: serine/threonine-rich
F;549-569/Domain: transmembrane #status predicted <TMM>
F;572-1049/Domain: intracellular #status predicted <INT>
F;840-869/Domain: desmoglein repeat <DG1>
F;870-899/Domain: desmoglein repeat <DG2>
F;900-927/Domain: desmoglein repeat <DG3>
F;928-956/Domain: desmoglein repeat <DG4>
F;969-1019/Region: glycine/serine-rich
F;110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 55; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 0.26;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDINI 15
:||||::: ||:|||
Db 97 YGIFVINQKTGEINI 111

RESULT 4
IJBOG1
desmoglein 1 precursor - bovine
N;Alternate names: desmoglein BDGM
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S14603; A38872; A37785; S38721; A48173; S24412
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
submitted to the EMBL Data Library, March 1991
A;Description: Complete sequence of the desmoglein precursor and evidence for the exi
A;Reference number: S14603
A;Accession: S14603
A;Molecule type: mRNA
A;Residues: 1-1043 <KOC>
A;Cross-references: EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti
A;Reference number: A38872; MUID:92037656
A;Accession: A38872
A;Molecule type: mRNA
A;Residues: 1-87;968-1043 <KO2>
A;Cross-references: GB:S64268; GB:S64270
R;Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990
A;Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion
A;Reference number: A37785; MUID:91097553
A;Accession: A37785
A;Molecule type: mRNA
A;Residues: 44-123,'V',125-493 <GOO>
A;Cross-references: GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
R;Zimbelmann, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S38721
A;Accession: S38721
A;Molecule type: mRNA
A;Residues: 44-1043 <ZIM>
A;Cross-references: EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
R;Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.
Eur. J. Cell Biol. 53, 1-12, 1990
A;Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m
A;Reference number: A48173; MUID:91168965
A;Accession: A48173
A;Molecule type: mRNA
A;Residues: 44-1001,'AQPPSAT' <KO3>

```
A;Cross-references: GB:X57784
A;Note: this sequence has been revised in references A38872 and S38721
C;Genetics:
A;Gene: DSG1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prc
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1043/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-491/Domain: cadherin repeat homology <CR4>
F;549-574/Domain: transmembrane #status predicted <TMM>
F;575-1043/Domain: intracellular #status predicted <INT>
F;846-875/Domain: desmoglein repeat <DG1>
F;876-905/Domain: desmoglein repeat <DG2>
F;906-933/Domain: desmoglein repeat <DG3>
F;934-962/Domain: desmoglein repeat <DG4>
F;963-1012/Region: glycine/serine-rich
F;110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             70.5%;  Score 55;  DB 1;  Length 1043;
  Best Local Similarity   60.0%;  Pred. No. 0.26;
  Matches    9;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDINI 15
           :|||:::  ||:|||
Db      97 YGIFVINQKTGEINI 111


RESULT   5
I50180
cadherin-7 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Aug-1999
C;Accession: I50180
R;Nakagawa, S.; Takeichi, M.
Development 121, 1321-1332, 1995
A;Title: Neural crest cell-cell adhesion controlled by sequential and subpopulation-spec
A;Reference number: I50178; MUID:95309115
A;Accession: I50180
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-785 <NAK>
A;Cross-references: GB:D42150; NID:g868000; PIDN:BAA07721.1; PID:g868001
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             65.4%;  Score 51;  DB 2;  Length 785;
  Best Local Similarity   66.7%;  Pred. No. 0.91;
  Matches    8;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           ||::|:|||||:
Db      94 IFIIDENTGDIH 105


RESULT   6
IJBODE
desmocollin 1a - bovine
N;Alternate names: desmocollin BDCM; desmosomal glycoprotein 2
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A43838; B43838; A38456; A60714; S14542
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Schmelz, M.; Franke, W.W.
Differentiation 47, 29-36, 1991
A;Title: Amino acid sequence of bovine muzzle epithelial desmocollin derived from cloned
A;Reference number: A43838; MUID:92008912
A;Accession: A43838
A;Molecule type: mRNA
A;Residues: 1-761 <KOC>
A;Cross-references: GB:X58029; GB:S57985; NID:g453583; PIDN:CAA41088.1; PID:g453584
A;Accession: B43838
A;Molecule type: protein
A;Residues: 1-32;65-76;148-159;164-176;190-205;208-219;238-256;361-375;377-388;478-48
A;Experimental source: muzzle epithelium
A;Note: sequence extracted from NCBI backbone
R;Collins, J.E.; Legan, P.K.; Kenny, T.P.; MacGarvie, J.; Holton, J.L.; Garrod, D.R.
J. Cell Biol. 113, 381-391, 1991
A;Title: Cloning and sequence analysis of desmosomal glycoproteins 2 and 3 (desmocoll
A;Reference number: A38456; MUID:91185414
A;Accession: A38456
A;Molecule type: mRNA
A;Residues: 606-761 <COL>
A;Cross-references: EMBL:X56967; NID:g310; PIDN:CAA40287.1; PID:g311
R;Holton, J.L.; Kenny, T.P.; Legan, P.K.; Collins, J.E.; Keen, J.N.; Sharma, R.; Garr
J. Cell Sci. 97, 239-246, 1990
A;Title: Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal similarity t
A;Reference number: A60714; MUID:91115997
A;Accession: A60714
A;Molecule type: protein
A;Residues: 1-6,'A',8-9,'R',11-17,'RCE',21-23 <HOL>
A;Experimental source: nasal epidermis
C;Genetics:
A;Gene: DSC1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycop
F;1-761/Product: desmocollin 1a #status experimental <MAT>
F;1-561/Domain: extracellular #status predicted <EXT>
F;3-108/Domain: cadherin repeat homology <CR1>
F;111-220/Domain: cadherin repeat homology <CR2>
F;223-338/Domain: cadherin repeat homology <CR3>
F;339-444/Domain: cadherin repeat homology <CR4>
F;445-561/Domain: cadherin repeat homology <CR5>
F;562-582/Domain: transmembrane #status predicted <TMM>
F;583-761/Domain: intracellular #status predicted <INT>
F;31,266,413/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;584,588,678/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status
F;605/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicte
F;671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicte
F;681/Binding site: phosphate (Tyr) (covalent) #status predicted
F;682/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicte


  Query Match             57.7%;  Score 45;  DB 1;  Length 761;
  Best Local Similarity   53.8%;  Pred. No. 8.9;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDI 13
           | :| ::|:||||
Db      48 FNLFFIEKDTGDI 60


RESULT   7
IJBODF
desmocollin 1b precursor - bovine
N;Alternate names: desmosomal glycoprotein 3
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: B38456; A39377; S14567
R;Collins, J.E.; Legan, P.K.; Kenny, T.P.; MacGarvie, J.; Holton, J.L.; Garrod, D.R.
J. Cell Biol. 113, 381-391, 1991
A;Title: Cloning and sequence analysis of desmosomal glycoproteins 2 and 3 (desmocoll
A;Reference number: A38456; MUID:91185414
A;Accession: B38456
A;Molecule type: mRNA
A;Residues: 1-839 <COL>
A;Cross-references: GB:X56966; NID:g315; PIDN:CAA40286.1; PID:g316
R;Mechanic, S.; Raynor, K.; Hill, J.E.; Cowin, P.
Proc. Natl. Acad. Sci. U.S.A. 88, 4476-4480, 1991
A;Title: Desmocollins form a distinct subset of the cadherin family of cell adhesion
A;Reference number: A39377; MUID:91239591
```

A;Accession: A39377
A;Molecule type: mRNA
A;Residues: 1-484,'A',486-839 <MEC>
A;Cross-references: GB:M67489; GB:M61750; NID:g162970; PIDN:AAA30492.1; PID:g162971
A;Note: part of this sequence, including the amino end of the mature protein, was confir
C;Genetics:
A;Gene: DSC1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprot
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-132/Domain: propeptide #status predicted <PRO>
F;133-839/Product: desmocollin 1b #status experimental <MAT>
F;133-682/Domain: extracellular #status predicted <EXT>
F;135-240/Domain: cadherin repeat homology <CR1>
F;243-352/Domain: cadherin repeat homology <CR2>
F;355-470/Domain: cadherin repeat homology <CR3>
F;471-576/Domain: cadherin repeat homology <CR4>
F;577-682/Domain: cadherin repeat homology <CR5>
F;694-714/Domain: transmembrane #status predicted <TMM>
F;718-839/Domain: intracellular #status predicted <INT>
F;163,398,545/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;716,720,810/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pre
F;737/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted
F;803,830/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predict
F;813/Binding site: phosphate (Tyr) (covalent) #status predicted
F;814/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 57.7%; Score 45; DB 1; Length 839;
Best Local Similarity 53.8%; Pred. No. 9.9;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDI 13
| :| ::|:|||
Db 180 FNLFFIEKDTGDI 192

RESULT 8
B48910
desmocollin 1b precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 15-Aug-1997
C;Accession: B48910
R;King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.
Genomics 18, 185-194, 1993
A;Title: Cloning of the cDNA (DSC1) coding for human type 1 desmocollin and its assignme
A;Reference number: A48910; MUID:94116981
A;Accession: B48910
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-770 <KIN>
A;Cross-references: GB:X72929
C;Genetics:
A;Gene: GDB:DSC1
A;Cross-references: GDB:128632
A;Map position: 18q12.2-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; glycoprotein; phosphop
F;67-172/Domain: cadherin repeat homology <CR1>

Query Match 57.7%; Score 45; DB 2; Length 770;
Best Local Similarity 53.8%; Pred. No. 9;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDI 13
| :| ::|:|||
Db 112 FNLFYIEKDTGDI 124

RESULT 9
A48910
desmocollin 1a precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 15-Aug-1997
C;Accession: A48910
R;King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.
Genomics 18, 185-194, 1993
A;Title: Cloning of the cDNA (DSC1) coding for human type 1 desmocollin and its assig
A;Reference number: A48910; MUID:94116981
A;Accession: A48910
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-824 <KIN>
A;Cross-references: GB:X72925
C;Genetics:
A;Gene: GDB:DSC1
A;Cross-references: GDB:128632
A;Map position: 18q12.2-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; glycoprotein; phosp
F;67-172/Domain: cadherin repeat homology <CR1>

Query Match 57.7%; Score 45; DB 2; Length 824;
Best Local Similarity 53.8%; Pred. No. 9.7;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDI 13
| :| ::|:|||
Db 112 FNLFYIEKDTGDI 124

RESULT 10
I37281
Dscla precursor - human
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 20-Aug-1999
C;Accession: I37281
R;Theis, D.G.; Koch, P.J.; Franke, W.W.
Int. J. Dev. Biol. 37, 101-110, 1993
A;Title: Differential synthesis of type 1 and type 2 desmocollin mRNAs in human strat
A;Reference number: I37281; MUID:93283249
A;Accession: I37281
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-840 <RES>
A;Cross-references: EMBL:Z34522; NID:g505536; PIDN:CAA84278.1; PID:g505537
C;Superfamily: cadherin; cadherin repeat homology

Query Match 57.7%; Score 45; DB 2; Length 840;
Best Local Similarity 53.8%; Pred. No. 9.9;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FGIFVVDKNTGDI 13
| :| ::|:|||
Db 182 FNLFYIEKDTGDI 194

RESULT 11
I37282
Dsc1b precursor - human
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 20-Aug-1999
C;Accession: I37282
R;Theis, D.G.; Koch, P.J.; Franke, W.W.
Int. J. Dev. Biol. 37, 101-110, 1993
A;Title: Differential synthesis of type 1 and type 2 desmocollin mRNAs in human strat
A;Reference number: I37281; MUID:93283249
A;Accession: I37282
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-894 <RES>
A;Cross-references: EMBL:Z34522; NID:g505536; PIDN:CAA84279.1; PID:g505538
C;Genetics:

```
A;Introns: 829/2
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             57.7%;  Score 45;  DB 2;  Length 894;
  Best Local Similarity   53.8%;  Pred. No. 11;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGIFVVDKNTGDI 13
           | :| ::|:||||
Db     182 FNLFYIEKDTGDI 194


RESULT  12
I37016
cadherin-6 - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 20-Aug-1999
C;Accession: I37016
R;Shimoyama, Y.; Gotoh, M.; Terasaki, T.; Kitajima, M.; Hirohashi, S.
Cancer Res. 55, 2206-2211, 1995
A;Title: Isolation and sequence analysis of human cadherin-6 complementary DNA for the f
A;Reference number: I37016; MUID:95262134
A;Accession: I37016
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-790 <RES>
A;Cross-references: GB:D31784; NID:g974184; PIDN:BAA06562.1; PID:g974185
C;Genetics:
A;Gene: GDB:CDH6
A;Cross-references: GDB:5822908
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             55.1%;  Score 43;  DB 2;  Length 790;
  Best Local Similarity   54.5%;  Pred. No. 20;
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT  13
I52701
K-cadherin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999
C;Accession: I52701
R;Xiang, Y.
Cancer Res. 54, 3034-3041, 1993
A;Title: Isolation of complementary DNA encoding K-cadherin, a novel rat cadherin prefer
A;Reference number: I52701; MUID:94243827
A;Accession: I52701
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-789 <RES>
A;Cross-references: GB:D25290; NID:g435460; PIDN:BAA04975.1; PID:g435461
C;Genetics:
A;Gene: KCAD
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             55.1%;  Score 43;  DB 2;  Length 789;
  Best Local Similarity   54.5%;  Pred. No. 20;
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT  14
I50178
cadherin-6B - chicken
C;Species: Gallus gallus (chicken)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Aug-1999
C;Accession: I50178
R;Nakagawa, S.; Takeichi, M.
Development 121, 1321-1332, 1995
A;Title: Neural crest cell-cell adhesion controlled by sequential and subpopulation-s
A;Reference number: I50178; MUID:95309115
A;Accession: I50178
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-790 <NAK>
A;Cross-references: GB:D42149; NID:g867998; PIDN:BAA07720.1; PID:g867999
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             55.1%;  Score 43;  DB 2;  Length 790;
  Best Local Similarity   54.5%;  Pred. No. 20;
  Matches    6;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDI 13
           :|::::|||||
Db     100 LFIINENTGDI 110


RESULT  15
G02678
cadherin-14 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C;Accession: G02678
R;Shibata, T.; Shimoyama, Y.; Gotoh, M.; Hirohashi, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: H01584
A;Accession: G02678
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-790 <SHI>
A;Cross-references: EMBL:U59325; NID:g1389852; PIDN:AAB02933.1; PID:g1389853
C;Superfamily: cadherin; cadherin repeat homology


  Query Match             55.1%;  Score 43;  DB 2;  Length 790;
  Best Local Similarity   58.3%;  Pred. No. 20;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 IFVVDKNTGDIN 14
           ||::|  ||||:
Db     100 IFIIDDTTGDIH 111


Search completed: January  7, 2000, 08:52:39
Job time: 2269 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:34 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-3
Perfect score: 71
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	71	100.0	999	1	R30742	Human pemphigus vu
2	71	100.0	614	1	W07908	Pemphigus vulgaris
3	71	100.0	15	1	W04843	Self epitope of de
4	71	100.0	15	1	W64815	Desmoglein-3 190-2
5	71	100.0	15	1	W78814	Desmoglein 3 prote
6	62	87.3	778	1	W15489	Pemphigus foliaceu
7	55	77.5	560	1	W13009	Segment of desmoso
8	55	77.5	263	1	W13010	Segment of desmoso
9	39	54.9	878	1	R55060	Sequence of human
10	39	54.9	878	1	R85487	Human E-cadherin p
11	34	47.9	4472	1	R97246	Virulence gene clu
12	34	47.9	2233	1	W48711	HPIV-3 JS isolate
13	34	47.9	2233	1	W48712	HPIV-3 FRhl cp45 v
14	34	47.9	2233	1	W48713	HPIV-3 Vero cp45 v
15	34	47.9	456	1	W82841	Human cerebral pro
16	34	47.9	330	1	W89745	Staphylococcus aur
17	33	46.5	2408	1	R24307	Translation of ORF
18	33	46.5	479	1	W22341	P. membranaefacien
19	33	46.5	81	1	W75861	Human secretory pr
20	33	46.5	438	1	W95500	B. subtilis YabE r
21	32	45.1	697	1	R58591	Hepatitis C virus
22	32	45.1	730	1	R82881	Lupin exo-(1-4)bet
23	32	45.1	532	1	R91035	Recombinant V8 pro
24	32	45.1	611	1	W02157	Periplasmic Beta-N
25	32	45.1	916	1	W13129	Full length human
26	32	45.1	205	1	W20360	H. pylori secreted
27	32	45.1	217	1	W20960	H. pylori secreted
28	32	45.1	537	1	W22220	Protein encoded by
29	32	45.1	645	1	W20054	S. aureus threonyl
30	32	45.1	532	1	W22219	Protein encoded by
31	32	45.1	916	1	W25658	Human cadherin-4.
32	32	45.1	481	1	W34554	M11TL glycosidase
33	32	45.1	419	1	W30917	DNA gyrase subunit
34	32	45.1	371	1	W41513	N. meningitidis al
35	32	45.1	371	1	W41514	N. gonorrheae alph
36	32	45.1	151	1	W52130	Putative insectici
37	32	45.1	277	1	W51011	Human liver carbon
38	32	45.1	276	1	W37921	Bos taurus DNase I
39	32	45.1	26	1	W60192	Bacteriophage spol
40	32	45.1	729	1	W60724	Human TK2 protein
41	32	45.1	398	1	W30523	Kurthia sp. KAPA s
42	32	45.1	645	1	W61633	Staphylococcus aur
43	32	45.1	481	1	W49858	Desulfurococcus MI
44	32	45.1	729	1	W79273	Human T2K (TRAF2-a
45	32	45.1	611	1	W85599	Hexosaminidase enz

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 71; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
||||||||||||||||
Db 190 LNSKIAFKIVSQEPA 204

RESULT 2
W07908
ID W07908 standard; protein; 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphis vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucuous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose

CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 71; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
||||||||||||||||
Db 189 LNSKIAFKIVSQEPA 203

RESULT 3
W04843
ID W04843 standard; peptide; 15 AA.
AC W04843;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN WO9627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1; Page 39; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-myslin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 190-204)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 71; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
||||||||||||||||
Db 1 LNSKIAFKIVSQEPA 15

RESULT 4
W64815
ID W64815 standard; peptide; 15 AA.
AC W64815;
DT 29-SEP-1998 (first entry)
DE Desmoglein-3 190-204.
KW Desmoglein; DG; gene therapy; pemphigus vulgaris; microparticle;
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PD 21-JUL-1998.
PF 22-JAN-1997; 787547.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS;
DR WPI; 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PT expression of proteins e.g. in gene therapy
PS Disclosure; Column 4; 42pp; English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA;

Query Match 100.0%; Score 71; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
||||||||||||||||
Db 1 LNSKIAFKIVSQEPA 15

RESULT 5
W78814
ID W78814 standard; peptide; 15 AA.
AC W78814;
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 190-204.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
OS Synthetic.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998; U01499.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR WPI; 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure; Page 8; 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less that 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (i) a
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment

```
CC   of a naturally-occurring protein from an infectious agent which infects
CC   a mammal; (2) a peptide having a length and sequence which permits it to
CC   bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC   peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC   to W78897 are peptide fragments for use in the present invention. The
CC   MPs are highly effective vehicles for the delivery of polynucleotides
CC   into phagocytic cells. They can be used for gene therapy, e.g. for
CC   treating genetic diseases, infections or tumours or for downregulating
CC   an immune response.
SQ   Sequence   15 AA;

  Query Match             100.0%;  Score 71;  DB 1;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          |||||||||||||||
Db      1 LNSKIAFKIVSQEPA 15


RESULT   6
W15489
ID   W15489 standard; Protein; 778 AA.
AC   W15489;
DT   17-JUN-1997  (first entry)
DE   Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW   Pemphigus foliaceus; autoantibody; constant region; IgG;
KW   extracellular region; antigen; hinge portion; skin;
KW   dermatitis herpetiformis; fusion protein; detection; ss.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1. .545
FT                   /note= "Pemphigus foliaceus antigen protein"
PN   J09077800-A.
PD   25-MAR-1997.
PF   12-SEP-1995; 260899.
PR   12-SEP-1995; JP-260899.
PA   (NISH/) NISHIKAWA T.
DR   WPI; 97-241758/22.
DR   P-PSDB; T66428.
PT   Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT   through the hinge region used to treat pemphigus foliaceus
PS   Claim 1; Page 10-12; 17pp; Japanese.
CC   This sequence represents a fused protein recognised by pemphigus
CC   foliaceus patient autoantibody which comprises the constant region
CC   of IgG linked to the extracellular region of pemphigus foliaceus
CC   antigen protein through the hinge portion.  Pemphigus foliaceus is
CC   a chronic, generalised, vesicular and scaling skin eruption similar
CC   to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC   protein is useful to treat pemphigus foliaceus.  The antigen is
CC   especially administered through an adsorbent upon which the fusion
CC   protein is immobilised via a carrier. The fusion protein is also
CC   useful for detecting pemphigus foliaceus antibodies which is useful
CC   in immunodiagnosis.  The fusion protein has little or no side effects.
SQ   Sequence   778 AA;

  Query Match             87.3%;  Score 62;  DB 1;  Length 778;
  Best Local Similarity   80.0%;  Pred. No. 0.00072;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          |||||||||: |||:
Db    190 LNSKIAFKIIRQEPS 204


RESULT   7
W13009
ID   W13009 standard; protein; 560 AA.
AC   W13009;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 7; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
CC   carcinomas.
SQ   Sequence   560 AA;

  Query Match             77.5%;  Score 55;  DB 1;  Length 560;
  Best Local Similarity   73.3%;  Pred. No. 0.011;
  Matches   11;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          |||||:::||| |||
Db    143 LNSKISYRIVSLEPA 157


RESULT   8
W13010
ID   W13010 standard; protein; 263 AA.
AC   W13010;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 9; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
```

```
CC   carcinomas.
SQ   Sequence   263 AA;


  Query Match             77.5%;  Score 55;  DB 1;  Length 263;
  Best Local Similarity   73.3%;  Pred. No. 0.0046;
  Matches   11;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEPA 15
           |||||:::||| |||
Db      41 LNSKISYRIVSLEPA 55


RESULT   9
R55060
ID   R55060 standard; Protein; 878 AA.
AC   R55060;
DT   08-NOV-1994  (first entry)
DE   Sequence of human liver E-cadherin.
KW   Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
KW   uvomorulin; L-CAM; Cell CAM 120/80.
OS   Homo sapiens.
PN   WO9411401-A.
PD   26-MAY-1994.
PF   16-NOV-1993; U11097.
PR   17-NOV-1992; US-978897.
PA   (UYYA ) UNIV YALE.
PI   Morrow JS,  Rimm DL;
DR   WPI; 94-183426/22.
DR   N-PSDB; Q65487.
PT   Purified human E-cadherin protein and nucleic acid - used to
PT   develop prods. for diagnosis, prognosis, therapy and prophylaxis
PT   of t-cadherin disorders, e.g. malignancies
PS   Claim 1; Page 59-63; 97pp; English.
CC   E-cadherin is a cell adhesion molecule that is also known as
CC   uvomorulin, L-CAM and Cell CAM 120/80. The DNA encoding hEC was obtd.
CC   by screening normal human liver and hepatocellular carcinoma cDNA
CC   libraries and a colonic epithelial cell cDNA library. The following
CC   sequences are specifically claimed: AAs 1-878; 151-878; 30
CC   sequential AAs from AAs 308-878; AAs 1-150; AAs 178-289; AAs 290-
CC   401; AAs 402-513; AAs 178-513; AAs 151-703; AAs 1-703; AAs 728-878;
CC   AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-
CC   2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-
CC   1648; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;
CC   1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from
CC   AAs 308-878. The prods. can be used in the diagnosis, prognosis,
CC   therapy and prophylaxis of conditions involving improper E-cadherin
CC   expression. Suitable dosages for i.v. admin. of a protein are
CC   20-500 mcg/kg body wt.
SQ   Sequence   878 AA;


  Query Match             54.9%;  Score 39;  DB 1;  Length 878;
  Best Local Similarity   53.8%;  Pred. No. 23;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |: ||: |:||:|
Db     293 NAAIAYTILSQDP 305


RESULT  10
R85487
ID   R85487 standard; Protein; 878 AA.
AC   R85487;
DT   18-MAR-1996  (first entry)
DE   Human E-cadherin precursor.
KW   E-cadherin; T-lymphocyte; alpha-E,beta-7 integrin; cell adhesion;
KW   autoimmune disease; Crohn disease; psoriasis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .150
FT                   /label= Sig_peptide
FT   domain          151. .702
FT                   /label= Extracellular_domain
FT                   /note= "the extracellular domain (amino acids 1-552
FT                   of the mature protein) is the preferred
FT                   region for generation of peptides of the
FT                   invention"
FT   domain          703. .726
FT                   /label= Transmembrane_domain
FT   domain          727. .876
FT                   /label= Cytoplasmic_domain
PN   WO9529693-A1.
PD   09-NOV-1995.
PF   03-MAY-1995; U05518.
PR   03-MAY-1994; US-237919.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PI   Brenner MB,  Cepek KL;
DR   WPI; 95-392921/50.
DR   N-PSDB; T05764.
PT   Inhibiting adhesion of T lymphocytes with E-cadherin - useful for
PT   isolating agents to treat auto-immune diseases e.g. Crohn's disease,
PT   psoriasis, etc
PS   Disclosure; Page 70-75; 103pp; English.
CC   The human E-cadherin protein precursor (R85487) is expressed by
CC   an cDNA clone (T05764) derived from human liver.  The extracellular
CC   domain of E-cadherin is used to generate peptides that specifically
CC   bind to heterotypic cognates of E-cadherin and which inhibit adhesion
CC   of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial
CC   or endothelial cells in vitro or in vivo, thereby modulating
CC   mucosal immune responses.  Such peptides are also specifically
CC   reactive with a monoclonal antibody (E4.6 or E6.1) that binds to
CC   E-cadherin and that can inhibit T-cell binding.
SQ   Sequence   878 AA;


  Query Match             54.9%;  Score 39;  DB 1;  Length 878;
  Best Local Similarity   53.8%;  Pred. No. 23;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |: ||: |:||:|
Db     293 NAAIAYTILSQDP 305


RESULT  11
R97246
ID   R97246 standard; Protein; 4472 AA.
AC   R97246;
DT   07-JAN-1997  (first entry)
DE   Virulence gene cluster polypeptide product.
KW   Mutant; adaptation; virulence factor; identification; screening;
KW   vaccine; drugs; infection; treatment.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   region          Region
FT                   /note= "All x's in this sequence correspond to
FT                   termination codons in the virulence gene
FT                   cluster sequence given in T09224."
PN   WO9617951-A2.
PD   13-JUN-1996.
PF   11-DEC-1995; G02875.
PR   09-DEC-1994; GB-024921.
PR   31-JAN-1995; GB-001881.
PR   05-MAY-1995; GB-009239.
PA   (RPMS-) RPMS TECHNOLOGY LTD.
PI   Holden DW;
DR   WPI; 96-287194/29.
PT   Identifying virulence genes in microorganisms - by introducing
PT   mutants with insertion inactivated genes into environment and
PT   retrieval and analysis of mutants
PS   Claim 51; Figure 41; 131pp; English.
CC   A method for identifying a microorganism having a reduced adaptation
CC   to a particular environment comprising the steps of: (1) providing a
CC   plurality of microorganisms each of which is independently mutated by
```

```
CC   the insertional inactivation of a gene with a nucleic acid comprising
CC   a unique marker sequence so that each mutant contains a different
CC   marker sequence, or clones of the said microorganism; (2) providing
CC   individually a stored sample of each mutant produced by step (1) and
CC   providing individually stored nucleic acid comprising the unique
CC   marker sequence from each individual mutant; (3) introducing a
CC   plurality of mutants produced by step (1) into the said particular
CC   environment and allowing those microorganisms which are able to do so
CC   to grow in the said environment; (4) retrieving microorganisms from
CC   the said environment or a selected part thereof and isolating the
CC   nucleic acid from the retrieved microorganisms; (5) comparing any
CC   marker sequences in the nucleic acid isolated in step (4) to the
CC   unique marker sequence of each individual mutant stored as in step
CC   (2); and (6) selecting an individual mutant which does not contain any
CC   of the marker sequences as isolated in step (4).  The products and
CC   methods can be used for identifying virulence genes in microorganisms.
CC   The mutant microorganisms can be used in vaccines or to screen for
CC   drugs which reduce virulence or compounds useful for preventing,
CC   ameliorating or treating infections in animals or plants.
SQ   Sequence   4472 AA;

  Query Match             47.9%;  Score 34;  DB 1;  Length 4472;
  Best Local Similarity   58.3%;  Pred. No. 1.3e+03;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQE 13
           |: |||::|| |
Db    2873 NNIIAFQVVSXE 2884


RESULT  12
W48711
ID   W48711 standard; Protein; 2233 AA.
AC   W48711;
DT   13-OCT-1998  (first entry)
DE   HPIV-3 JS isolate wild-type L protein.
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
OS   Human parainfluenza virus.
PN   WO9813501-A2.
PD   02-APR-1998.
PF   19-SEP-1997; U16718.
PR   27-SEP-1996; US-026823.
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR   WPI; 98-230710/20.
DR   N-PSDB; V18272.
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
PS   Disclosure; Page 246-254; 426pp; English.
CC   This sequence represents the wild-type L protein from Human parainfluenza
CC   virus (HPIV-3) type 3 isolate JS. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
SQ   Sequence   2233 AA;

  Query Match             47.9%;  Score 34;  DB 1;  Length 2233;
  Best Local Similarity   35.7%;  Pred. No. 6.1e+02;
  Matches    5;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
           |:  : ::|::|||
Db     973 LDRSVLYRIMNQEP 986


RESULT  13
W48712
ID   W48712 standard; Protein; 2233 AA.
AC   W48712;
DT   13-OCT-1998  (first entry)
DE   HPIV-3 FRhl cp45 vaccine L protein.
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
OS   Human parainfluenza virus.
PN   WO9813501-A2.
PD   02-APR-1998.
PF   19-SEP-1997; U16718.
PR   27-SEP-1996; US-026823.
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR   WPI; 98-230710/20.
DR   N-PSDB; V18273\.
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
PS   Disclosure; Page 265-273; 426pp; English.
CC   This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC   vaccine FRhl cp45 L protein. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
SQ   Sequence   2233 AA;

  Query Match             47.9%;  Score 34;  DB 1;  Length 2233;
  Best Local Similarity   35.7%;  Pred. No. 6.1e+02;
  Matches    5;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
           |:  : ::|::|||
Db     973 LDRSVLYRIMNQEP 986


RESULT  14
W48713
ID   W48713 standard; Protein; 2233 AA.
AC   W48713;
DT   13-OCT-1998  (first entry)
DE   HPIV-3 Vero cp45 vaccine L protein.
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
OS   Human parainfluenza virus.
PN   WO9813501-A2.
PD   02-APR-1998.
PF   19-SEP-1997; U16718.
PR   27-SEP-1996; US-026823.
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR   WPI; 98-230710/20.
DR   N-PSDB; V18274.
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
PS   Disclosure; Page 283-291; 426pp; English.
CC   This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC   vaccine Vero cp45 L protein. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
```

```
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
SQ   Sequence   2233 AA;


  Query Match             47.9%;  Score 34;  DB 1;  Length 2233;
  Best Local Similarity   35.7%;  Pred. No. 6.1e+02;
  Matches    5;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
           |:' : ::|::|||
Db     973 LDRSVLYRIMNQEP 986


RESULT  15
W82841
ID   W82841 standard; Protein; 456 AA.
AC   W82841;
DT   03-FEB-1999  (first entry)
DE   Human cerebral protein-1.
KW   Human; cerebral protein-1; HUCEP-1; ischaemic brain disease;
KW   neuron function activating activity; nerve denatured disease;
KW   Parkinson's disease; Alzheimer's disease.
OS   Homo sapiens.
PN   J10257891-A.
PD   29-SEP-1998.
PF   19-MAR-1997; 065716.
PR   19-MAR-1997; JP-065716.
PA   (TAIS ) TAISHO PHARM CO LTD.
DR   WPI; 98-575902/49.
DR   N-PSDB; V64162.
PT   HUCEP-1 protein having neuron function activating activity - useful
PT   for treating ischaemic brain diseases and nerve denaturation
PT   conditions such as Parkinson's and Alzheimer's diseases
PS   Claim 1; Fig 4; 17pp; Japanese.
CC   The present sequence represents human cerebral protein-1 (HUCEP-1).
CC   HUCEP-1 has neuron function activating activity. HUCEP-1 is useful
CC   for treating ischaemic brain diseases and nerve denaturation
CC   conditions such as Parkinson's and Alzheimer's diseases.
SQ   Sequence   456 AA;


  Query Match             47.9%;  Score 34;  DB 1;  Length 456;
  Best Local Similarity   54.5%;  Pred. No. 98;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SKIAFKIVSQE 13
           :|: |||||::
Db     408 TKLGFKIVSKD 418


Search completed: January  6, 2000, 15:40:36
Job time: 1683 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:39 ; Search time 65.16 Seconds
(without alignments)
10.858 Million cell updates/sec

Title: US-08-991-628-3
Perfect score: 71
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database : PIR_62:*

Word size : 0

Number of hits that pass the threshold : 142080
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	71	100.0	999	1	IJHUG3	desmoglein 3 precu
2	62	87.3	1049	1	IJHUG1	desmoglein 1 precu
3	62	87.3	1043	1	IJBOG1	desmoglein 1 precu
4	55	77.5	1117	2	S38673	desmoglein 2 - hum
5	42	59.2	884	1	IJMSCE	E-cadherin precurs
6	42	59.2	884	2	S34438	uvomorulin - mouse
7	41	57.7	732	1	IJCHCB	B-cadherin precurs
8	40	56.3	5147	1	IJFFTM	cadherin-related t
9	40	56.3	1421	2	T02501	hypothetical prote
10	39	54.9	906	1	IJHUCN	cadherin 2 precurs
11	39	54.9	882	1	IJHUCE	cadherin 1 precurs
12	39	54.9	726	1	S73915	virulence-associat
13	39	54.9	829	2	I46536	Ksp-cadherin - rab
14	38	53.5	906	1	IJXLC2	N-cadherin 2 precu
15	38	53.5	905	1	IJXLC1	N-cadherin 1 precu
16	38	53.5	905	2	S43064	cadherin - African
17	38	53.5	385	2	S49752	homeotic protein Y
18	38	53.5	138	2	S74520	hypothetical prote
19	38	53.5	533	2	JC5040	positive regulator
20	38	53.5	533	2	S57904	virR49 protein - S
21	37	52.1	493	1	S39532	aldehyde dehydroge
22	37	52.1	236	2	D71733	phosphoribosylamin
23	37	52.1	461	2	A71662	sodium/pantothenat
24	37	52.1	135	2	H72471	hypothetical prote
25	36	50.7	877	1	IJBOCN	N-cadherin precurs
26	36	50.7	906	1	IJMSCN	N-cadherin precurs
27	36	50.7	887	1	IJCHCL	E-cadherin precurs
28	36	50.7	829	1	IJHUCP	cadherin 3 precurs
29	36	50.7	822	1	IJMSCP	P-cadherin precurs
30	36	50.7	725	1	E64211	virulence-associat
31	36	50.7	276	2	JC5285	carbonyl reductase
32	36	50.7	277	2	JC5284	carbonyl reductase
33	36	50.7	475	2	H71660	nitrogen assimilat
34	36	50.7	742	2	A49341	isocitrate dehydro
35	36	50.7	103	2	D70897	hypothetical prote
36	36	50.7	182	2	H64386	hypothetical prote
37	36	50.7	1264	2	S64146	probable membrane
38	36	50.7	2122	2	B75009	ribonucleotide red
39	35	49.3	342	1	CFPM	plastoquinol--plas
40	35	49.3	320	1	CFNT	plastoquinol--plas
41	35	49.3	320	1	CFRZ	plastoquinol--plas
42	35	49.3	643	1	S24382	nitrous-oxide redu
43	35	49.3	2228	1	ZLNZSV	genome polyprotein
44	35	49.3	2048	1	ZLNZSE	genome polyprotein
45	35	49.3	241	2	S63634	ribosomal protein

ALIGNMENTS

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N;Alternate names: pemphigus vulgaris antigen
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41088
R;Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 869-877, 1991
A;Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A;Reference number: A41088; MUID:92069753
A;Accession: A41088
A;Molecule type: mRNA
A;Residues: 1-999 <AMA>
A;Cross-references: GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190752
C;Genetics:
A;Gene: GDB:DSG3
A;Cross-references: GDB:134030; OMIM:169615
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-999/Product: desmoglein homolog #status predicted <MAT>
F;50-615/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-267/Domain: cadherin repeat homology <CR2>
F;270-383/Domain: cadherin repeat homology <CR3>
F;390-495/Domain: cadherin repeat homology <CR4>
F;496-598/Domain: cadherin repeat homology <CR5>
F;616-639/Domain: transmembrane #status predicted <TMM>
F;640-999/Domain: intracellular #status predicted <INT>
F;910-938/Domain: desmoglein repeat <DG1>
F;937-966/Domain: desmoglein repeat <DG2>
F;110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 71; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 LNSKIAFKIVSQEPA 15
          |||||||||||||||
Db    190 LNSKIAFKIVSQEPA 204
```

RESULT 2
IJHUG1
desmoglein 1 precursor - human
N;Alternate names: desmosomal glycoprotein I
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S16906; A39706; A61254; A61279; S16158
R;Buxton, R.S.
submitted to the EMBL Data Library, November 1990
A;Reference number: S16906
A;Accession: S16906
A;Molecule type: mRNA

A;Residues: 1-1049 <BUX>
A;Cross-references: EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
R;Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; R
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991
A;Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,
A;Reference number: A39706; MUID:91271279
A;Accession: A39706
A;Molecule type: mRNA
A;Residues: 24-1049 <WHE>
A;Cross-references: GB:X56654
R;Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
J. Cell Sci. 99, 809-821, 1991
A;Title: Structural analysis and expression of human desmoglein: a cadherin-like compone
A;Reference number: A61254; MUID:92121251
A;Accession: A61254
A;Molecule type: mRNA
A;Residues: 26-1049 <NIL>
R;Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee
Biochem. Soc. Trans. 19, 1060-1064, 1991
A;Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfami
A;Reference number: A61279; MUID:92175187
A;Accession: A61279
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-55 <WH3>
C;Genetics:
A;Gene: GDB:DSG1
A;Cross-references: GDB:126563; OMIM:125670
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1049/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-493/Domain: cadherin repeat homology <CR4>
F;509-530/Region: serine/threonine-rich
F;549-569/Domain: transmembrane #status predicted <TMM>
F;572-1049/Domain: intracellular #status predicted <INT>
F;840-869/Domain: desmoglein repeat <DG1>
F;870-899/Domain: desmoglein repeat <DG2>
F;900-927/Domain: desmoglein repeat <DG3>
F;928-956/Domain: desmoglein repeat <DG4>
F;969-1019/Region: glycine/serine-rich
F;110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.3%; Score 62; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 0.0014;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
|||||||||: |||:
Db 190 LNSKIAFKIIRQEPS 204

RESULT 3
IJBOG1
desmoglein 1 precursor - bovine
N;Alternate names: desmoglein BDGM
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S14603; A38872; A37785; S38721; A48173; S24412
R;Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
submitted to the EMBL Data Library, March 1991
A;Description: Complete sequence of the desmoglein precursor and evidence for the existe
A;Reference number: S14603
A;Accession: S14603
A;Molecule type: mRNA
A;Residues: 1-1043 <KOC>
A;Cross-references: EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti
A;Reference number: A38872; MUID:92037656
A;Accession: A38872
A;Molecule type: mRNA
A;Residues: 1-87;968-1043 <KO2>
A;Cross-references: GB:S64268; GB:S64270
R;Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990
A;Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion
A;Reference number: A37785; MUID:91097553
A;Accession: A37785
A;Molecule type: mRNA
A;Residues: 44-123,'V',125-493 <GOO>
A;Cross-references: GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
R;Zimbelmann, R.
submitted to the EMBL Data Library, February 1991
A;Reference number: S38721
A;Accession: S38721
A;Molecule type: mRNA
A;Residues: 44-1043 <ZIM>
A;Cross-references: EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
R;Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.
Eur. J. Cell Biol. 53, 1-12, 1990
A;Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m
A;Reference number: A48173; MUID:91168965
A;Accession: A48173
A;Molecule type: mRNA
A;Residues: 44-1001,'AQPPSAT' <KO3>
A;Cross-references: GB:X57784
A;Note: this sequence has been revised in references A38872 and S38721
C;Genetics:
A;Gene: DSG1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-49/Domain: propeptide #status predicted <PRO>
F;50-1043/Product: desmoglein #status predicted <MAT>
F;50-548/Domain: extracellular #status predicted <EXT>
F;52-157/Domain: cadherin repeat homology <CR1>
F;160-269/Domain: cadherin repeat homology <CR2>
F;272-385/Domain: cadherin repeat homology <CR3>
F;392-491/Domain: cadherin repeat homology <CR4>
F;549-574/Domain: transmembrane #status predicted <TMM>
F;575-1043/Domain: intracellular #status predicted <INT>
F;846-875/Domain: desmoglein repeat <DG1>
F;876-905/Domain: desmoglein repeat <DG2>
F;906-933/Domain: desmoglein repeat <DG3>
F;934-962/Domain: desmoglein repeat <DG4>
F;963-1012/Region: glycine/serine-rich
F;110/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.3%; Score 62; DB 1; Length 1043;
Best Local Similarity 80.0%; Pred. No. 0.0014;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
|||||||||: |||:
Db 190 LNSKIAFKIIRQEPS 204

RESULT 4
S38673
desmoglein 2 - human
N;Alternate names: desmoglein HDGC
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C;Accession: S38673; B38872
R;Zimbelmann, R.

submitted to the EMBL Data Library, September 1993
A;Reference number: S38673
A;Accession: S38673
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1117 <ZIM>
A;Cross-references: EMBL:Z26317; NID:g416177; PIDN:CAA81226.1; PID:g416178
R;Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
Eur. J. Cell Biol. 55, 200-208, 1991
A;Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide
A;Reference number: A38872; MUID:92037656
A;Accession: B38872
A;Molecule type: mRNA
A;Residues: 777-1117 <KOC>
A;Cross-references: GB:S64273
C;Genetics:
A;Gene: GDB:DSG2
A;Cross-references: GDB:128808; OMIM:125671
A;Map position: 18q12.1-18q12.2
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein
F;51-158/Domain: cadherin repeat homology <CR1>
F;161-271/Domain: cadherin repeat homology <CR2>

Query Match 77.5%; Score 55; DB 2; Length 1117;
Best Local Similarity 73.3%; Pred. No. 0.032;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
|||||:::||| |||
Db 191 LNSKISYRIVSLEPA 205

RESULT 5
IJMSCE
E-cadherin precursor, epithelial - mouse
N;Alternate names: uvomorulin
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S04528; S03160; I49565; S48735
R;Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.; Takeichi, M.
Nature 329, 341-343, 1987
A;Title: Transformation of cell adhesion properties by exogenously introduced E-cadherin
A;Reference number: S04528; MUID:87315445
A;Accession: S04528
A;Molecule type: mRNA
A;Residues: 1-412,'V',414-884 <NAG>
A;Cross-references: EMBL:X06115
R;Ringwald, M.; Schuh, R.; Vestweber, D.; Eistetter, H.; Lottspeich, F.; Engel, J.; Doel
EMBO J. 6, 3647-3653, 1987
A;Title: The structure of cell adhesion molecule uvomorulin. Insights into the molecular
A;Reference number: S03160; MUID:88111553
A;Accession: S03160
A;Molecule type: mRNA
A;Residues: 157-884 <RIN>
A;Cross-references: EMBL:X06339
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Behrens, J.; Loewrick, O.; Klein-Hitpass, L.; Birchmeier, W.
Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991
A;Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithe
A;Reference number: I49565; MUID:92107977
A;Accession: I49565
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-15 <RES>
A;Cross-references: GB:M81449; NID:g192325; PIDN:AAA37352.1; PID:g192326
R;Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Porumb, T.; Takeichi, M.; Ikura, M.
FEBS Lett. 352, 318-322, 1994
A;Title: Purification and spectroscopic characterization of a recombinant amino-terminal
A;Reference number: S48735; MUID:95010732
A;Accession: S48735
A;Status: preliminary
A;Molecule type: protein
A;Residues: 156-300 <TON>
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though
C;Genetics:
A;Gene: E-cadherin
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-156/Domain: propeptide #status predicted <PRO>
F;157-884/Product: E-cadherin, epithelial #status experimental <MAT>
F;157-699/Domain: extracellular #status predicted <EXT>
F;159-264/Domain: cadherin repeat homology <CR1>
F;234-239/Region: cadherin binding #status predicted
F;267-377/Domain: cadherin repeat homology <CR2>
F;380-488/Domain: cadherin repeat homology <CR3>
F;489-597/Domain: cadherin repeat homology <CR4>
F;598-702/Domain: cadherin repeat homology <CR5>
F;702-733/Domain: transmembrane #status predicted <TMM>
F;734-884/Domain: intracellular #status predicted <INT>
F;842-855/Region: serine-rich
F;560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 59.2%; Score 42; DB 1; Length 884;
Best Local Similarity 61.5%; Pred. No. 7.1;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NSKIAFKIVSQEP 14
|: ||: ||||:|
Db 299 NAAIAYTIVSQDP 311

RESULT 6
S34438
uvomorulin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997
C;Accession: S34438
R;Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.
Nucleic Acids Res. 19, 6533-6539, 1991
A;Title: The structure of the gene coding for the mouse cell adhesion molecule uvomor
A;Reference number: S34438; MUID:92093614
A;Accession: S34438
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-884 <RIN>
A;Cross-references: EMBL:X60975
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
C;Genetics:
A;Introns: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1;
C;Superfamily: cadherin; cadherin repeat homology
F;380-488/Domain: cadherin repeat homology <CR3>

Query Match 59.2%; Score 42; DB 2; Length 884;
Best Local Similarity 61.5%; Pred. No. 7.1;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NSKIAFKIVSQEP 14
|: ||: ||||:|
Db 299 NAAIAYTIVSQDP 311

RESULT 7
IJCHCB
B-cadherin precursor - chicken (fragment)
N;Alternate names: K-CAM protein
C;Species: Gallus gallus (chicken)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A41634; A38715; S16160
R;Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
A;Title: Genes for two calcium-dependent cell adhesion molecules have similar structu

A;Reference number: A41634; MUID:92107987
A;Accession: A41634
A;Molecule type: DNA
A;Residues: 1-732 <SOR>
A;Cross-references: GB:M81894; NID:g212226; PIDN:AAA48929.1; PID:g212227
R;Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
J. Cell Biol. 113, 893-905, 1991
A;Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
A;Reference number: A38715; MUID:91225083
A;Accession: A38715
A;Molecule type: mRNA
A;Residues: 7-413,'V',415-732 <NAP>
A;Cross-references: GB:X58518; NID:g63113; PIDN:CAA41408.1; PID:g63114
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t
C;Genetics:
A;Gene: K-CAM
A;Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 293/3; 375/2; 423/1; 498/1; 571/1; 614/3; 66
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-6/Domain: propeptide (fragment) #status predicted <PRO>
F;6-554/Domain: extracellular #status predicted <EXT>
F;7-732/Product: B-cadherin #status predicted <MAT>
F;9-114/Domain: cadherin repeat homology <CR1>
F;84-89/Region: cadherin binding #status predicted
F;117-227/Domain: cadherin repeat homology <CR2>
F;230-339/Domain: cadherin repeat homology <CR3>
F;340-447/Domain: cadherin repeat homology <CR4>
F;448-552/Domain: cadherin repeat homology <CR5>
F;555-580/Domain: transmembrane #status predicted <TMM>
F;581-732/Domain: intracellular #status predicted <INT>
F;689-702/Region: serine-rich
F;137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 57.7%; Score 41; DB 1; Length 732;
Best Local Similarity 61.5%; Pred. No. 8.9;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 NSKIAFKIVSQEP 14
           |  ||: |:||||
Db     149 NGVIAYSILSQEP 161
```

RESULT 8
IJFFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C;Accession: A41087; B41087
R;Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A;Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A;Reference number: A41087; MUID:92069752
A;Accession: A41087
A;Molecule type: mRNA
A;Residues: 143-485;1279-5147 <MAH>
A;Cross-references: GB:M80537
A;Accession: B41087
A;Molecule type: DNA
A;Residues: 1-142;487-1278 <MA2>
A;Cross-references: GB:M80537
A;Note: 1229-Gly and 1233-Ser were also found
C;Genetics:
A;Gene: fat
A;Cross-references: FlyBase:FBgn0001075
C;Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C;Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
F;36-4583/Domain: extracellular #status predicted <EXT>
F;51-156/Domain: cadherin repeat homology <CR1>
F;159-270/Domain: cadherin repeat homology <CR2>
F;271-382/Domain: cadherin repeat homology <CR3>
F;390-494/Domain: cadherin repeat homology <CR4>
F;497-599/Domain: cadherin repeat homology <CR5>
F;602-708/Domain: cadherin repeat homology <CR6>
F;718-822/Domain: cadherin repeat homology <CR7>
F;831-942/Domain: cadherin repeat homology <CR8>
F;948-1049/Domain: cadherin repeat homology <CR9>
F;1052-1153/Domain: cadherin repeat homology <C10>
F;1156-1278/Domain: cadherin repeat homology <C11>
F;1281-1384/Domain: cadherin repeat homology <C12>
F;1387-1489/Domain: cadherin repeat homology <C13>
F;1492-1601/Domain: cadherin repeat homology <C14>
F;1607-1713/Domain: cadherin repeat homology <C15>
F;1717-1823/Domain: cadherin repeat homology <C16>
F;1826-1922/Domain: cadherin repeat homology <C17>
F;1925-2027/Domain: cadherin repeat homology <C18>
F;2028-2167/Domain: cadherin repeat homology <C19>
F;2169-2278/Domain: cadherin repeat homology <C20>
F;2281-2384/Domain: cadherin repeat homology <C99>
F;2387-2491/Domain: cadherin repeat homology <C21>
F;2494-2596/Domain: cadherin repeat homology <C22>
F;2599-2703/Domain: cadherin repeat homology <C23>
F;2707-2810/Domain: cadherin repeat homology <C24>
F;2813-2913/Domain: cadherin repeat homology <C25>
F;2915-3013/Domain: cadherin repeat homology <C26>
F;3014-3124/Domain: cadherin repeat homology <C27>
F;3127-3229/Domain: cadherin repeat homology <C28>
F;3232-3334/Domain: cadherin repeat homology <C29>
F;3337-3439/Domain: cadherin repeat homology <C30>
F;3442-3545/Domain: cadherin repeat homology <C31>
F;3548-3651/Domain: cadherin repeat homology <C32>
F;3654-3756/Domain: cadherin repeat homology <C33>
F;3954-4010/Domain: EGF homology <EG1>
F;4017-4048/Domain: EGF homology <EG2>
F;4056-4089/Domain: EGF homology <EG3>
F;4096-4127/Domain: EGF homology <EG4>
F;4584-4609/Domain: transmembrane #status predicted <TMM>
F;4610-5147/Domain: intracellular #status predicted <INT>

Query Match 56.3%; Score 40; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 LNSKIAFKIVSQEP 14
           || |::: |  |||
Db    1520 LNGKVSYAISKQEP 1533
```

RESULT 9
T02501
hypothetical protein T19C21.7 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 14-May-1999
C;Accession: T02501
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.
A;Reference number: Z14676
A;Accession: T02501
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1421 <ROU>
A;Cross-references: EMBL:AC004683; NID:g3395421; PID:g3395428
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
A;Introns: 52/2; 107/1; 148/2; 191/3; 212/2; 1317/3; 1346/3
A;Note: T19C21.7

Query Match 56.3%; Score 40; DB 2; Length 1421;
Best Local Similarity 42.9%; Pred. No. 29;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 LNSKIAFKIVSQEP 14
          ::|||:::|::| |
Db    262 VDSKISYEIITQNP 275


RESULT  10
IJHUCN
cadherin 2 precursor - human
N;Alternate names: N-cadherin; neuronal cadherin
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A38870; S11487; JQ0751; S13799
R;Reid, R.A.
submitted to the EMBL Data Library, November 1990
A;Reference number: A38870
A;Accession: A38870
A;Molecule type: mRNA
A;Residues: 1-906 <REI>
A;Cross-references: EMBL:X54315; NID:g34998; PIDN:CAA38213.1; PID:g34999
R;Reid, R.A.; Hemperly, J.J.
Nucleic Acids Res. 18, 5896, 1990
A;Title: Human N-cadherin: nucleotide and deduced amino acid sequence.
A;Reference number: S11487; MUID:91016946
A;Accession: S11487
A;Molecule type: mRNA
A;Residues: 1-340,'N',342-698,'R',700-704,'F',706-906 <RE2>
A;Cross-references: EMBL:X54315
A;Note: this sequence has been revised in reference A38870
R;Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow,
J. Neurochem. 55, 805-812, 1990
A;Title: N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin
A;Reference number: JQ0751; MUID:90347462
A;Accession: JQ0751
A;Molecule type: mRNA
A;Residues: 160-194,'IR',197-211,'L',213-227,'Q',229,'N',231-235,'G',237-248,'T',250-356
A;Cross-references: GB:M34064
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH2; NCAD
A;Cross-references: GDB:128185; OMIM:114020
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-159/Domain: propeptide #status predicted <PRO>
F;160-906/Product: N-cadherin #status predicted <MAT>
F;160-714/Domain: extracellular #status predicted <EXT>
F;162-267/Domain: cadherin repeat homology <CR1>
F;237-242/Region: cadherin binding #status predicted
F;270-382/Domain: cadherin repeat homology <CR2>
F;385-497/Domain: cadherin repeat homology <CR3>
F;500-605/Domain: cadherin repeat homology <CR4>
F;606-712/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TMM>
F;747-906/Domain: intracellular #status predicted <INT>
F;865-878/Region: serine-rich
F;190,273,325,402,572,622,651,692/Binding site: carbohydrate (Asn) (covalent) #status pr


  Query Match             54.9%;  Score 39;  DB 1;  Length 906;
  Best Local Similarity   46.7%;  Pred. No. 27;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          ||  : ::|||| |:
Db    300 LNGMLRYRIVSQAPS 314


RESULT  11
IJHUCE
cadherin 1 precursor - human
N;Alternate names: ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomo
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S37654; S31430; S05475; S31460; S06716; A57171; JC2230; I52294; I52704;
R;Bussemakers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.
Mol. Biol. Rep. 17, 123-128, 1993
A;Title: Molecular cloning and characterization of the human E-cadherin cDNA.
A;Reference number: S37654; MUID:93211394
A;Accession: S37654
A;Molecule type: mRNA
A;Residues: 1-882 <BUS>
A;Cross-references: EMBL:Z13009; NID:g31072; PIDN:CAA78353.1; PID:g31073
R;Kelker, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.
submitted to the EMBL Data Library, December 1992
A;Description: Sequence of human E-cadherin cDNA.
A;Reference number: S31430
A;Accession: S31430
A;Molecule type: mRNA
A;Residues: 1-542,'F',544-882 <KEL>
A;Cross-references: EMBL:Z18923; NID:g31074; PIDN:CAA79356.1; PID:g31075
R;Mansouri, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.
Differentiation 38, 67-71, 1988
A;Title: Characterization and chromosomal localization of the gene encoding the human
A;Reference number: S05475; MUID:89031725
A;Accession: S05475
A;Molecule type: mRNA
A;Residues: 157-311 <MAN>
A;Cross-references: EMBL:X12790
A;Note: nucleotide sequence is not complete
R;Frixen, U.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S31460
A;Accession: S31460
A;Molecule type: mRNA
A;Residues: 265-392 <FRI>
A;Cross-references: EMBL:X52279; NID:g28821; PIDN:CAA36522.1; PID:g28822
R;Wheelock, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.
J. Cell. Biochem. 34, 187-202, 1987
A;Title: Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.
A;Reference number: S06716; MUID:87280410
A;Accession: S06716
A;Molecule type: protein
A;Residues: 'XQ',157-162,'V',164-179 <WHE>
R;Berx, G.; Staes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.G.; van Bokhov
Genomics 26, 281-289, 1995
A;Title: Cloning and characterization of the human invasion suppressor gene E-cadheri
A;Reference number: A57171; MUID:95324920
A;Accession: A57171
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-30,32-882 <BER>
A;Cross-references: GB:L34784
R;Rimm, D.L.; Morrow, J.S.
Biochem. Biophys. Res. Commun. 200, 1754-1761, 1994
A;Title: Molecular cloning of human E-cadherin suggests a novel subdivision of the c
A;Reference number: JC2230; MUID:94242050
A;Accession: JC2230
A;Molecule type: mRNA
A;Residues: 1-9,'G',11-15,'RSPLGSQERSPPPCLTRELHVHGAPAPPEKRPR',52-67,'I',69,'LTPIP',7:
A;Cross-references: GB:L08599; NID:g340184; PIDN:AAA61259.1; PID:g340185
A;Note: the majority of differences between this and other reports represent apparan
A;Note: the authors translated the codon CCG for residue 868 as Arg
R;Bussemakers, M.J.G.; Giroldi, L.A.; van Bokhoven, A.; Schalken, J.A.
Biochem. Biophys. Res. Commun. 203, 1284-1290, 1994
A;Title: Transcriptional regulation of the human E-cadherin gene in human prostate c
A;Reference number: I52294; MUID:94380041
A;Accession: I52294
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-16 <RES>
A;Cross-references: GB:L34545; NID:g509604; PIDN:AAA21764.1; PID:g509605
R;Becker, K.F.; Atkinson, M.J.; Reich, U.; Becker, I.; Nekarda, H.; Siewert, J.R.; H
Cancer Res. 54, 3845-3852, 1994
A;Title: E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.
```

```
A;Reference number: I52704; MUID:94306394
A;Accession: I52704
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 337-476 <RE2>
A;Cross-references: GB:S72492; NID:g632756
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH1; UVO
A;Cross-references: GDB:120484; OMIM:192090
A;Map position: 16q22.1-16q22.1
A;Introns: 379/3; 440/3
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-154/Domain: propeptide #status predicted <PRO>
F;155-882/Product: E-cadherin #status experimental <MAT>
F;155-697/Domain: extracellular #status predicted <EXT>
F;157-262/Domain: cadherin repeat homology <CR1>
F;232-237/Region: cadherin binding #status predicted
F;265-375/Domain: cadherin repeat homology <CR2>
F;378-486/Domain: cadherin repeat homology <CR3>
F;487-595/Domain: cadherin repeat homology <CR4>
F;596-700/Domain: cadherin repeat homology <CR5>
F;698-731/Domain: transmembrane #status predicted <TMM>
F;732-882/Domain: intracellular #status predicted <INT>
F;840-853/Region: serine-rich
F;376,558,570,622,637,849/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             54.9%;  Score 39;  DB 1;  Length 882;
  Best Local Similarity   53.8%;  Pred. No. 26;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEP 14
          |: ||: |:||:|
Db    297 NAAIAYTILSQDP 309


RESULT 12
S73915
virulence-associated protein vacB homolog - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
N;Alternate names: hypothetical protein K04_orf726
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S73915
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885
A;Accession: S73915
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-726 <HIM>
A;Cross-references: EMBL:AE000057; GB:U00089; NID:g1674279; PIDN:AAB96237.1; PID:g167429
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: vacB
A;Genetic code: SGC3
C;Superfamily: virulence-associated protein vacB homolog

  Query Match             54.9%;  Score 39;  DB 1;  Length 726;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEPA 15
          |:|:||:|   :||
Db    193 NTKVAFQIDEYDPA 206


RESULT 13
I46536
Ksp-cadherin - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 14-Feb-1997
C;Accession: I46536
R;Thomson, R.B.; Igarashi, P.; Biemesderfer, D.; Kim, R.; Abu-Alfa, A.; Soleimani, M.
J. Biol. Chem. 270, 17594-17601, 1995
A;Title: Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific member o
A;Reference number: I46536; MUID:95340560
A;Accession: I46536
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-829 <THO>
A;Cross-references: EMBL:U28945; NID:g902885; PID:g902886

  Query Match             54.9%;  Score 39;  DB 2;  Length 829;
  Best Local Similarity   57.1%;  Pred. No. 24;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEPA 15
          || : | |:|| ||
Db    160 NSDLRFHILSQTPA 173


RESULT 14
IJXLC2
N-cadherin 2 precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A43785
R;Ginsberg, D.; DeSimone, D.; Geiger, B.
Development 111, 315-325, 1991
A;Title: Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early
A;Reference number: A43785; MUID:91372132
A;Accession: A43785
A;Molecule type: mRNA
A;Residues: 1-906 <GIN>
A;Cross-references: GB:X57675; NID:g64910; PIDN:CAA40867.1; PID:g64911
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-160/Domain: propeptide #status predicted <PRO>
F;161-906/Product: N-cadherin 2 #status predicted <MAT>
F;161-714/Domain: extracellular #status predicted <EXT>
F;163-268/Domain: cadherin repeat homology <CR1>
F;238-243/Region: cadherin binding #status predicted
F;271-383/Domain: cadherin repeat homology <CR2>
F;386-498/Domain: cadherin repeat homology <CR3>
F;501-606/Domain: cadherin repeat homology <CR4>
F;607-714/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TMM>
F;747-906/Domain: intracellular #status predicted <INT>
F;863-878/Region: serine-rich
F;274,326,403,573,623,652,693/Binding site: carbohydrate (Asn) (covalent) #status pre

  Query Match             53.5%;  Score 38;  DB 1;  Length 906;
  Best Local Similarity   50.0%;  Pred. No. 42;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEPA 15
          |  : :||:|| ||
Db    302 NGMLRYKILSQTPA 315


RESULT 15
IJXLC1
N-cadherin 1 precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 19-Jan-1996
C;Accession: JQ0442
```

R;Detrick, R.J.; Dickey, D.; Kintner, C.R.
Neuron 4, 493-506, 1990
A;Title: The effects of N-cadherin misexpression on morphogenesis in Xenopus embryos.
A;Reference number: JQ0442; MUID:90211966
A;Accession: JQ0442
A;Molecule type: mRNA
A;Residues: 1-905 <DET>
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prc
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-160/Domain: propeptide #status predicted <PRO>
F;161-905/Product: N-cadherin 1 #status predicted <MAT>
F;161-713/Domain: extracellular #status predicted <EXT>
F;163-268/Domain: cadherin repeat homology <CR1>
F;238-243/Region: cadherin binding #status predicted
F;271-383/Domain: cadherin repeat homology <CR2>
F;386-498/Domain: cadherin repeat homology <CR3>
F;501-606/Domain: cadherin repeat homology <CR4>
F;607-713/Domain: cadherin repeat homology <CR5>
F;714-745/Domain: transmembrane #status predicted <TMM>
F;746-905/Domain: intracellular #status predicted <INT>
F;862-877/Region: serine-rich
F;191,274,326,403,573,623,651,692/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 53.5%; Score 38; DB 1; Length 905;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2 NSKIAFKIVSQEPA 15
           |  : :||:|| ||
Db     302 NGMLRYKILSQTPA 315
```

Search completed: January 7, 2000, 08:52:41
Job time: 2271 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:37 ; Search time 28.55 Seconds
(without alignments)
15.282 Million cell updates/sec

Title: US-08-991-628-3
Perfect score: 71
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database : SwissProt_38:*

Word size : 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	71	100.0	999	1	DSG3_HUMAN	P32926 homo sapien
2	62	87.3	1043	1	DSG1_BOVIN	Q03763 bos taurus
3	62	87.3	1049	1	DSG1_HUMAN	Q02413 homo sapien
4	55	77.5	1117	1	DSG2_HUMAN	Q14126 homo sapien
5	42	59.2	884	1	CAD1_MOUSE	P09803 mus musculu
6	41	57.7	732	1	CADB_CHICK	P33145 gallus gall
7	40	56.3	5147	1	FAT_DROME	P33450 drosophila
8	39	54.9	882	1	CAD1_HUMAN	P12830 homo sapien
9	39	54.9	906	1	CAD2_HUMAN	P19022 homo sapien
10	39	54.9	726	1	VACB_MYCPN	P75529 mycoplasma
11	38	53.5	905	1	CADB_XENLA	P33152 xenopus lae
12	38	53.5	905	1	CADN_XENLA	P20310 xenopus lae
13	38	53.5	906	1	CADO_XENLA	P33147 xenopus lae
14	38	53.5	385	1	YOX1_YEAST	P34161 saccharomyc
15	37	52.1	493	1	DHAG_HUMAN	P49189 homo sapien
16	36	50.7	887	1	CAD1_CHICK	P08641 gallus gall
17	36	50.7	877	1	CAD2_BOVIN	P19534 bos taurus
18	36	50.7	906	1	CAD2_MOUSE	P15116 mus musculu
19	36	50.7	829	1	CAD3_HUMAN	P22223 homo sapien
20	36	50.7	822	1	CAD3_MOUSE	P10287 mus musculu
21	36	50.7	276	1	DHCA_RAT	P47727 rattus norv
22	36	50.7	741	1	IDH2_VIBA1	P41561 vibrio sp.
23	36	50.7	725	1	VACB_MYCGE	P47350 mycoplasma
24	36	50.7	182	1	Y696_METJA	Q58107 methanococc
25	36	50.7	1264	1	YGN3_YEAST	P53125 saccharomyc
26	35	49.3	1146	1	AS10_YEAST	P48361 saccharomyc
27	35	49.3	320	1	CYF_MAIZE	P46617 zea mays (m
28	35	49.3	318	1	CYF_OENHO	P04658 oenothera h
29	35	49.3	320	1	CYF_ORYSA	P07888 oryza sativ
30	35	49.3	320	1	CYF_PEA	P00155 pisum sativ
31	35	49.3	320	1	CYF_PORPU	P51265 porphyra pu
32	35	49.3	320	1	CYF_TOBAC	P06449 nicotiana t
33	35	49.3	320	1	CYF_VICFA	P06669 vicia faba
34	35	49.3	320	1	CYF_WHEAT	P05151 triticum ae
35	35	49.3	1419	1	MDR_PLAFF	P13568 plasmodium
36	35	49.3	643	1	NOSZ_ALCEU	Q59105 alcaligenes
37	35	49.3	2228	1	RRPL_SEND5	P27566 sendai viru
38	35	49.3	2048	1	RRPL_SENDE	P06829 sendai viru
39	35	49.3	2228	1	RRPL_SENDF	Q06996 sendai viru
40	35	49.3	2228	1	RRPL_SENDZ	P06447 sendai viru
41	35	49.3	809	1	TER1_CAEEL	P54811 caenorhabdi
42	35	49.3	760	1	VACB_SHIFL	P30851 shigella fl
43	34	47.9	541	1	ASNH_METJA	Q58516 methanococc
44	34	47.9	633	1	BEM4_YEAST	P39011 saccharomyc
45	34	47.9	367	1	YSNK_CAEEL	Q19417 caenorhabdi

ALIGNMENTS

RESULT 1
DSG3_HUMAN
ID DSG3_HUMAN STANDARD; PRT; 999 AA.
AC P32926;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92069753.
RA AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion.";
RL Cell 67:869-877(1991).
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC ----------

CC ----------
DR EMBL; M76482; AAA60230.1; -.
DR PIR; A41088; IJHUG3.
DR HSSP; P09803; 1EDH.
DR MIM; 169615; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.

FT	SIGNAL	1 23	POTENTIAL.
FT	PROPEP	24 49	POTENTIAL.
FT	CHAIN	50 999	DESMOGLEIN 3.
FT	DOMAIN	50 615	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	616 640	POTENTIAL.
FT	DOMAIN	641 999	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50 158	CADHERIN 1.
FT	REPEAT	159 268	CADHERIN 2.
FT	REPEAT	269 383	CADHERIN 3.
FT	REPEAT	386 499	CADHERIN 4.
FT	REPEAT	910 935	DESMOGLEIN REPEAT 1.
FT	REPEAT	936 966	DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110 110	POTENTIAL.

```
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    459    459       POTENTIAL.
FT   CARBOHYD    545    545       POTENTIAL.
SQ   SEQUENCE   999 AA;  107503 MW;  4891F6AE CRC32;


  Query Match             100.0%;  Score 71;  DB 1;  Length 999;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          ||||||||||||||||
Db    190 LNSKIAFKIVSQEPA 204


RESULT   2
DSG1_BOVIN
ID   DSG1_BOVIN     STANDARD;      PRT;  1043 AA.
AC   Q03763;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
GN   DSG1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
OC   Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RA   KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANKE W.W.;
RL   Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 44-1043 FROM N.A.
RC   TISSUE=MUZZLE EPITHELIUM;
RX   MEDLINE; 91168965.
RA   KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA   ZIMBELMANN R., FRANKE W.W.;
RT   "Identification of desmoglein, a constitutive desmosomal
RT   glycoprotein, as a member of the cadherin family of cell adhesion
RT   molecules.";
RL   Eur. J. Cell Biol. 53:1-12(1990).
RN   [3]
RP   REVISIONS, AND SEQUENCE OF 101-123.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA   FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
RN   [4]
RP   SEQUENCE OF 44-493 FROM N.A.
RX   MEDLINE; 91097553.
RA   GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT   "Desmoglein shows extensive homology to the cadherin family of cell
RT   adhesion molecules.";
RL   Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X58466; CAA41380.1; -.
DR   EMBL; X57784; CAA40930.1; -.
DR   EMBL; M58165; AAA62709.1; -.
DR   PIR; S14603; IJBOG1.
DR   HSSP; P09803; 1EDH.
DR   PFAM; PF00028; cadherin; 3.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1043       DESMOGLEIN 1.
FT   DOMAIN       50    548       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    549    573       POTENTIAL.
FT   DOMAIN      574   1043       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    498       CADHERIN 4.
FT   REPEAT      819    845       DESMOGLEIN REPEAT 1.
FT   REPEAT      846    875       DESMOGLEIN REPEAT 2.
FT   REPEAT      876    905       DESMOGLEIN REPEAT 3.
FT   REPEAT      906    933       DESMOGLEIN REPEAT 4.
FT   REPEAT      934    962       DESMOGLEIN REPEAT 5.
FT   DOMAIN      963   1012       GLY/SER-RICH.
FT   CARBOHYD    110    110
FT   CARBOHYD    180    180       POTENTIAL.
FT   CARBOHYD    496    496       POTENTIAL.
FT   CONFLICT    124    124       I -> V (IN REF. 4).
SQ   SEQUENCE   1043 AA;  112243 MW;  13898584 CRC32;


  Query Match             87.3%;  Score 62;  DB 1;  Length 1043;
  Best Local Similarity   80.0%;  Pred. No. 0.00067;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          |||||||||: |||:
Db    190 LNSKIAFKIIRQEPS 204


RESULT   3
DSG1_HUMAN
ID   DSG1_HUMAN     STANDARD;      PRT;  1049 AA.
AC   Q02413;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN   DSG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE; 91271279.
RA   WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOTIS P., POYNTER D.,
RA   ARNEMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA   BUXTON R.S., MAGEE A.I.;
RT   "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT   junctions, is related to the cadherin family of cell adhesion
RT   molecules.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND ESOPHAGUS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
```

```
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56654; CAA39976.1; -.
DR   PIR; S16906; IJHUG1.
DR   HSSP; P09803; 1EDH.
DR   MIM; 125670; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     49       POTENTIAL.
FT   CHAIN        50   1049       DESMOGLEIN 1.
FT   DOMAIN       50    545       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    546    570       POTENTIAL.
FT   DOMAIN      571   1049       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       50    158       CADHERIN 1.
FT   REPEAT      159    270       CADHERIN 2.
FT   REPEAT      271    385       CADHERIN 3.
FT   REPEAT      386    497       CADHERIN 4.
FT   REPEAT      813    839       DESMOGLEIN REPEAT 1.
FT   REPEAT      840    869       DESMOGLEIN REPEAT 2.
FT   REPEAT      870    899       DESMOGLEIN REPEAT 3.
FT   REPEAT      900    927       DESMOGLEIN REPEAT 4.
FT   REPEAT      928    956       DESMOGLEIN REPEAT 5.
FT   DOMAIN      969   1019       GLY/SER-RICH.
FT   CARBOHYD     36     36       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    180    180       POTENTIAL.
SQ   SEQUENCE   1049 AA;  113715 MW;  FDD79961 CRC32;


  Query Match             87.3%;  Score 62;  DB 1;  Length 1049;
  Best Local Similarity   80.0%;  Pred. No. 0.00067;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEPA 15
           |||||||||: |||:
Db     190 LNSKIAFKIIRQEPS 204


RESULT   4
DSG2_HUMAN
ID   DSG2_HUMAN     STANDARD;      PRT;  1117 AA.
AC   Q14126;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DESMOGLEIN 2 PRECURSOR (HDGC).
GN   DSG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON CARCINOMA;
RX   MEDLINE; 94192736.
RA   SCHAEFER S., KOCH P.J., FRANKE W.W.;
RT   "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT   expression catalogue of the desmoglein subfamily of desmosomal
RT   cadherins.";
RL   Exp. Cell Res. 211:391-399(1994).
RN   [2]
RP   SEQUENCE OF 777-1117 FROM N.A.
RX   MEDLINE; 92037656.
RA   KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANKE W.W.;
RT   "Complete amino acid sequence of the epidermal desmoglein precursor
RT   polypeptide and identification of a second type of desmoglein gene.";
RL   Eur. J. Cell Biol. 55:200-208(1991).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC       INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC       FILAMENTS MEDIATING CELL-CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z26317; CAA81226.1; -.
DR   HSSP; P15116; 1NCI.
DR   MIM; 125671; -.
DR   PFAM; PF00028; cadherin; 4.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Cytoskeleton; Calcium-binding.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     48       POTENTIAL.
FT   CHAIN        49   1117       DESMOGLEIN 2.
FT   DOMAIN       49    608       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    609    633       POTENTIAL.
FT   DOMAIN      634   1117       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       49    159       CADHERIN 1.
FT   REPEAT      160    272       CADHERIN 2.
FT   REPEAT      273    387       CADHERIN 3.
FT   REPEAT      388    502       CADHERIN 4.
FT   REPEAT      880    911       DESMOGLEIN REPEAT 1.
FT   REPEAT      912    941       DESMOGLEIN REPEAT 2.
FT   REPEAT      942    967       DESMOGLEIN REPEAT 3.
FT   REPEAT      968    991       DESMOGLEIN REPEAT 4.
FT   REPEAT      992   1020       DESMOGLEIN REPEAT 5.
FT   REPEAT     1021   1050       DESMOGLEIN REPEAT 6.
FT   CARBOHYD    111    111       POTENTIAL.
FT   CARBOHYD    181    181       POTENTIAL.
FT   CARBOHYD    308    308       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    513    513       POTENTIAL.
SQ   SEQUENCE   1117 AA;  122385 MW;  84D3B898 CRC32;


  Query Match             77.5%;  Score 55;  DB 1;  Length 1117;
  Best Local Similarity   73.3%;  Pred. No. 0.015;
  Matches   11;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEPA 15
           |||||:::|||  |||
Db     191 LNSKISYRIVSLEPA 205


RESULT   5
CAD1_MOUSE
ID   CAD1_MOUSE     STANDARD;      PRT;   884 AA.
AC   P09803; Q61377;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UVOMORULIN) (ARC-1).
GN   CDH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE; 87315445.
RA   NAGAFUCHI A., SHIRAYOSHI Y., OKAZARI K., YASUDA K., TAKEICHI M.;
RT   "Transformation of cell adhesion properties by exogenously introduced
RT   E-cadherin cDNA.";
RL   Nature 329:341-343(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE; 92093614.
RA   RINGWALD M., BARIBAULT H., SCHMIDT C., KEMLER R.;
RT   "The structure of the gene coding for the mouse cell adhesion
RT   molecule uvomorulin.";
RL   Nucleic Acids Res. 19:6533-6539(1991).
RN   [3]
RP   SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.
RX   MEDLINE; 88111553.
RA   RINGWALD M., SCHUH R., VESTWEBER D., EISTETTER H., LOTTSPEICH F.,
RA   ENGEL J., DOELZ R., JAEHNIG F., EPPLEN J., MAYER S., MUELLER C.,
RA   KEMLER R.;
RT   "The structure of cell adhesion molecule uvomorulin. Insights into
RT   the molecular mechanism of Ca2+-dependent cell adhesion.";
RL   EMBO J. 6:3647-3653(1987).
RN   [4]
RP   SEQUENCE OF 1-15 FROM N.A.
RX   MEDLINE; 92107977.
RA   BEHRENS J., LOEWRICK O., KLEIN-HITPASS L., BIRCHMEIER W.;
RT   "The E-cadherin promoter: functional analysis of a G.C-rich region
RT   and an epithelial cell-specific palindromic regulatory element.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:11495-11499(1991).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
RX   MEDLINE; 96176249.
RA   NAGAR B., OVERDUIN M., IKURA M., RINI J.M.;
RT   "Structural basis of calcium-induced E-cadherin rigidification and
RT   dimerization.";
RL   Nature 380:360-364(1996).
RN   [6]
RP   STRUCTURE BY NMR OF 157-260.
RX   MEDLINE; 96271285.
RA   OVERDUIN M., TONG K.I., KAY C.M., IKURA M.;
RT   "1H, 15N and 13C resonance assignments and monomeric structure of the
RT   amino-terminal extracellular domain of epithelial cadherin.";
RL   J. Biomol. NMR 7:173-189(1996).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06115; CAA29488.1; -.
DR   EMBL; X60961; CAA43292.1; -.
DR   EMBL; X60962; CAA43292.1; JOINED.
DR   EMBL; X60963; CAA43292.1; JOINED.
DR   EMBL; X60964; CAA43292.1; JOINED.
DR   EMBL; X60965; CAA43292.1; JOINED.
DR   EMBL; X60966; CAA43292.1; JOINED.
DR   EMBL; X60967; CAA43292.1; JOINED.
DR   EMBL; X60968; CAA43292.1; JOINED.
DR   EMBL; X60969; CAA43292.1; JOINED.
DR   EMBL; X60970; CAA43292.1; JOINED.
DR   EMBL; X60971; CAA43292.1; JOINED.
DR   EMBL; X60972; CAA43292.1; JOINED.
DR   EMBL; X60973; CAA43292.1; JOINED.
DR   EMBL; X60974; CAA43292.1; JOINED.
DR   EMBL; X60975; CAA43292.1; JOINED.
DR   EMBL; X06339; CAA29645.1; -.
DR   EMBL; M81449; AAA37352.1; -.
DR   PIR; S04528; IJMSCE.
DR   PDB; 1EDH; 11-JAN-97.
DR   PDB; 1SUH; 11-JUL-96.
DR   MGD; MGI:88354; CDH1.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal; 3D-structure.
FT   SIGNAL       1    27       POTENTIAL.
FT   PROPEP      28   156       POTENTIAL.
FT   CHAIN      157   884       EPITHELIAL-CADHERIN.
FT   DOMAIN     157   709       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   710   733       POTENTIAL.
FT   DOMAIN     734   884       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     157   264       CADHERIN 1.
FT   REPEAT     265   377       CADHERIN 2.
FT   REPEAT     378   488       CADHERIN 3.
FT   REPEAT     489   595       CADHERIN 4.
FT   REPEAT     596   699       CADHERIN 5.
FT   DOMAIN     840   855       SER-RICH.
FT   CARBOHYD   560   560       POTENTIAL.
FT   CARBOHYD   639   639       POTENTIAL.
FT   CONFLICT   267   267       E -> P (IN REF. 2).
FT   CONFLICT   272   272       S -> F (IN REF. 2).
SQ   SEQUENCE   884 AA;  98255 MW;  8C758910 CRC32;


  Query Match           59.2%;  Score 42;  DB 1;  Length 884;
  Best Local Similarity  61.5%;  Pred. No. 3.6;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |: ||: ||||:|
Db     299 NAAIAYTIVSQDP 311


RESULT   6
CADB_CHICK
ID   CADB_CHICK     STANDARD;      PRT;   732 AA.
AC   P33145;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   B-CADHERIN PRECURSOR (K-CAM PROTEIN) (FRAGMENT).
GN   K-CAM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92107987.
RA   SORKIN B.C., GALLIN W.J., EDELMAN G.M., CUNNINGHAM B.A.;
RT   "Genes for two calcium-dependent cell adhesion molecules have similar
RT   structures and are arranged in tandem in the chicken genome.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).
RN   [2]
RP   SEQUENCE OF 7-732 FROM N.A.
RC   TISSUE=EMBRYONIC BRAIN;
RX   MEDLINE; 91225083.
RA   NAPOLITANO E.W., VENSTROM K., WHEELER E.F., REICHARDT L.F.;
RT   "Molecular cloning and characterization of B-cadherin, a novel chick
RT   cadherin.";
RL   J. Cell Biol. 113:893-905(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
```

```
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. B-CADHERIN MAY HAVE IMPORTANT
CC       FUNCTIONS IN NEUROGENESIS, IN AT LEAST SOME EPITHELIA, AND IN
CC       EMBRYOGENESIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81894; AAA48929.1; -.
DR   EMBL; X58518; CAA41408.1; -.
DR   PIR; A41634; IJCHCB.
DR   HSSP; P09803; 1SUH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat.
FT   NON_TER       1      1
FT   PROPEP       <1      6       POTENTIAL.
FT   CHAIN         7    732       B-CADHERIN.
FT   DOMAIN        6    554       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    555    580       POTENTIAL.
FT   DOMAIN      581    732       CYTOPLASMIC (POTENTIAL).
FT   REPEAT        6    114       CADHERIN 1.
FT   REPEAT      115    227       CADHERIN 2.
FT   REPEAT      228    339       CADHERIN 3.
FT   REPEAT      340    443       CADHERIN 4.
FT   REPEAT      444    554       CADHERIN 5.
FT   DOMAIN      689    702       SER-RICH.
FT   CARBOHYD    137    137       POTENTIAL.
FT   CARBOHYD    410    410       POTENTIAL.
FT   CONFLICT    414    414       M -> V (IN REF. 2).
SQ   SEQUENCE   732 AA;  80613 MW;  9C3CF5E9 CRC32;


  Query Match             57.7%;  Score 41;  DB 1;  Length 732;
  Best Local Similarity   61.5%;  Pred. No. 4.5;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |  ||: |:|||
Db     149 NGVIAYSILSQEP 161


RESULT   7
FAT_DROME
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN   FT.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92069752.
RA   MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA   GOODMAN C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   Cell 67:853-868(1991).
CC   -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC       TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC       DIFFERENTIATION AND MORPHOGENESIS, AND DEAT DURING THE PUPAL
CC       STAGE.
CC   -!- MISCELLANEOUS: SIMILARITY; CONTAINS 37 CADHERINS-TYPE REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80537; AAA28530.1; -.
DR   PIR; A41087; IJFFTM.
DR   HSSP; P00740; 1IXA.
DR   FLYBASE; FBgn0001075; ft.
DR   PFAM; PF00008; EGF; 4.
DR   PFAM; PF00028; cadherin; 34.
DR   PFAM; PF00054; laminin_G; 2.
DR   PROSITE; PS00232; CADHERIN; 22.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
KW   Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW   Calcium-binding; Repeat; EGF-like domain.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN       36   4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4584   4609       POTENTIAL.
FT   DOMAIN     4610   5147       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       36    156       CADHERIN 1.
FT   REPEAT      157    270       CADHERIN 2.
FT   REPEAT      271    382       CADHERIN 3.
FT   REPEAT      383    494       CADHERIN 4.
FT   REPEAT      495    599       CADHERIN 5.
FT   REPEAT      600    708       CADHERIN 6.
FT   REPEAT      709    820       CADHERIN 7.
FT   REPEAT      821    942       CADHERIN 8.
FT   REPEAT      943   1049       CADHERIN 9.
FT   REPEAT     1050   1153       CADHERIN 10.
FT   REPEAT     1154   1278       CADHERIN 11.
FT   REPEAT     1279   1384       CADHERIN 12.
FT   REPEAT     1385   1489       CADHERIN 13.
FT   REPEAT     1490   1601       CADHERIN 14.
FT   REPEAT     1602   1713       CADHERIN 15.
FT   REPEAT     1714   1823       CADHERIN 16.
FT   REPEAT     1824   1922       CADHERIN 17.
FT   REPEAT     1923   2027       CADHERIN 18.
FT   REPEAT     2028   2167       CADHERIN 19.
FT   REPEAT     2168   2278       CADHERIN 20.
FT   REPEAT     2279   2385       CADHERIN 21.
FT   REPEAT     2386   2491       CADHERIN 22.
FT   REPEAT     2492   2596       CADHERIN 23.
FT   REPEAT     2597   2703       CADHERIN 24.
FT   REPEAT     2704   2810       CADHERIN 25.
FT   REPEAT     2811   2913       CADHERIN 26.
FT   REPEAT     2914   3013       CADHERIN 27.
FT   REPEAT     3014   3124       CADHERIN 28.
FT   REPEAT     3125   3229       CADHERIN 29.
FT   REPEAT     3230   3334       CADHERIN 30.
FT   REPEAT     3335   3439       CADHERIN 31.
FT   REPEAT     3440   3545       CADHERIN 32.
FT   REPEAT     3546   3651       CADHERIN 33.
FT   REPEAT     3652   3756       CADHERIN 34.
FT   DOMAIN     3950   4011       EGF-LIKE 1.
FT   DOMAIN     4013   4049       EGF-LIKE 2.
```

```
FT   DOMAIN     4052   4090       EGF-LIKE 3.
FT   DOMAIN     4092   4128       EGF-LIKE 4.
FT   DOMAIN     4321   4362       EGF-LIKE 5.
FT   DISULFID   3954   3966       BY SIMILARITY.
FT   DISULFID   3960   3999       BY SIMILARITY.
FT   DISULFID   4001   4010       BY SIMILARITY.
FT   DISULFID   4017   4028       BY SIMILARITY.
FT   DISULFID   4022   4037       BY SIMILARITY.
FT   DISULFID   4039   4048       BY SIMILARITY.
FT   DISULFID   4056   4067       BY SIMILARITY.
FT   DISULFID   4061   4078       BY SIMILARITY.
FT   DISULFID   4080   4089       BY SIMILARITY.
FT   DISULFID   4096   4107       BY SIMILARITY.
FT   DISULFID   4101   4116       BY SIMILARITY.
FT   DISULFID   4118   4127       BY SIMILARITY.
FT   DISULFID   4325   4341       BY SIMILARITY.
FT   DISULFID   4334   4350       BY SIMILARITY.
FT   DISULFID   4352   4361       BY SIMILARITY.
FT   CARBOHYD    239    239       POTENTIAL.
FT   CARBOHYD    257    257       POTENTIAL.
FT   CARBOHYD    276    276       POTENTIAL.
FT   CARBOHYD    280    280       POTENTIAL.
FT   CARBOHYD    402    402       POTENTIAL.
FT   CARBOHYD    461    461       POTENTIAL.
FT   CARBOHYD    605    605       POTENTIAL.
FT   CARBOHYD    631    631       POTENTIAL.
FT   CARBOHYD   1155   1155       POTENTIAL.
FT   CARBOHYD   1367   1367       POTENTIAL.
FT   CARBOHYD   1458   1458       POTENTIAL.
FT   CARBOHYD   1751   1751       POTENTIAL.
FT   CARBOHYD   1831   1831       POTENTIAL.
FT   CARBOHYD   1880   1880       POTENTIAL.
FT   CARBOHYD   2080   2080       POTENTIAL.
FT   CARBOHYD   2171   2171       POTENTIAL.
FT   CARBOHYD   2247   2247       POTENTIAL.
FT   CARBOHYD   2290   2290       POTENTIAL.
FT   CARBOHYD   2437   2437       POTENTIAL.
FT   CARBOHYD   2581   2581       POTENTIAL.
FT   CARBOHYD   2799   2799       POTENTIAL.
FT   CARBOHYD   2920   2920       POTENTIAL.
FT   CARBOHYD   2946   2946       POTENTIAL.
FT   CARBOHYD   2967   2967       POTENTIAL.
FT   CARBOHYD   3167   3167       POTENTIAL.
FT   CARBOHYD   3303   3303       POTENTIAL.
FT   CARBOHYD   3386   3386       POTENTIAL.
FT   CARBOHYD   3389   3389       POTENTIAL.
FT   CARBOHYD   3525   3525       POTENTIAL.
FT   CARBOHYD   3852   3852       POTENTIAL.
FT   CARBOHYD   3865   3865       POTENTIAL.
FT   CARBOHYD   3905   3905       POTENTIAL.
FT   CARBOHYD   4306   4306       POTENTIAL.
FT   CARBOHYD   4414   4414       POTENTIAL.
FT   CARBOHYD   4471   4471       POTENTIAL.
FT   CARBOHYD   4487   4487       POTENTIAL.
FT   CARBOHYD   4539   4539       POTENTIAL.
FT   CARBOHYD   4550   4550       POTENTIAL.
FT   VARIANT    1229   1229       S -> G.
FT   VARIANT    1233   1233       G -> S.
SQ   SEQUENCE   5147 AA;  564868 MW;  1EF20E13 CRC32;


  Query Match             56.3%;  Score 40;  DB 1;  Length 5147;
  Best Local Similarity   50.0%;  Pred. No. 54;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
           || |::: |  |||
Db    1520 LNGKVSYAISKQEP 1533


RESULT   8
CAD1_HUMAN
ID   CAD1_HUMAN     STANDARD;      PRT;   882 AA.
AC   P12830; Q14216;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UVOMORULIN) (CAM 120/80).
GN   CDH1 OR UVO OR CDHE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93211394.
RA   BUSSEMAKERS M.J.G., MEES S.G.M., VAN BOKHOVEN A., DEBRUYNE F.M.J.,
RA   SCHALKEN J.A.;
RT   "Molecular cloning and characterization of the human E-cadherin
RT   cDNA.";
RL   Mol. Biol. Rep. 17:123-128(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   KELKER W., WARDA A., ODA T., HIROHASHI S., KEMLER R., BIRCHMEIER W.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 172-311 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 89031725.
RA   MANSOURI A., SPURR N., GOODFELLOW P.N., KEMLER R.;
RT   "Characterization and chromosomal localization of the gene encoding
RT   the human cell adhesion molecule uvomorulin.";
RL   Differentiation 38:67-71(1988).
RN   [4]
RP   SEQUENCE OF 1-16 FROM N.A.
RX   MEDLINE; 94380041.
RA   BUSSEMAKERS M.J., GIROLDI L.A., VAN BOKHOVEN A., SCHALKEN J.A.;
RT   "Transcriptional regulation of the human E-cadherin gene in human
RT   prostate cancer cell lines: characterization of the human E-cadherin
RT   gene promoter.";
RL   Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
RN   [5]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 98415721.
RA   BERX G., BECKER K.-F., HOEFLER H., VAN ROY F.;
RT   "Mutations of the human E-cadherin (CDH1) gene.";
RL   Hum. Mutat. 12:226-237(1998).
RN   [6]
RP   VARIANTS ALA-370 AND ASN-473.
RX   MEDLINE; 94306394.
RA   BECKER K.-F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
RA   SIEWERT J.R., HOEFLER H.;
RT   "E-cadherin gene mutations provide clues to diffuse type gastric
RT   carcinomas.";
RL   Cancer Res. 54:3845-3852(1994).
RN   [7]
RP   VARIANT LOBULAR BREAST CARCINOMA SER-315.
RX   MEDLINE; 95049851.
RA   KANAI Y., ODA T., TSUDA H., OCHIAI A., HIROHASHI S.;
RT   "Point mutation of the E-cadherin gene in invasive lobular carcinoma
RT   of the breast.";
RL   Jpn. J. Cancer Res. 85:1035-1039(1994).
RN   [8]
RP   VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
RX   MEDLINE; 94355985.
RA   RISINGER J.I., BERCHUCK A., KOHLER M.F., BOYD J.;
RT   "Mutations of the E-cadherin gene in human gynecologic cancers.";
RL   Nat. Genet. 7:98-102(1994).
RN   [9]
RP   VARIANT GASTRIC ADENOCARCINOMA GLY-274--PRO-277 DEL.
RX   MEDLINE; 94173928.
RA   ODA T., KANAI Y., OYAMA T., YOSHIURA K., SHIMOYAMA Y., BIRCHMEIER W.,
RA   SUGIMURA T., HIROHASHI S.;
RT   "E-cadherin gene mutations in human gastric carcinoma cell lines.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
RN   [10]
RP   VARIANT GASTRIC CARCINOMA PRO-193.
```

```
RX   MEDLINE; 96390918.
RA   MUTA H., NOGUCHI M., KANAI Y., OCHIAI A., NAWATA H., HIROHASHI S.;
RT   "E-cadherin gene mutations in signet ring cell carcinoma of the
RT   stomach.";
RL   Jpn. J. Cancer Res. 87:843-848(1996).
RN   [11]
RP   VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND ASP-418--PHE-423 DEL.
RX   MEDLINE; 97197648.
RA   TAMURA G., SAKATA K., NISHIZUKA S., MAESAWA C., SUZUKI Y., IWAYA T.,
RA   TERASHIMA M., SAITO K., SATODATE R.;
RT   "Inactivation of the E-cadherin gene in primary gastric carcinomas and
RT   gastric carcinoma cell lines.";
RL   Jpn. J. Cancer Res. 87:1153-1159(1996).
RN   [12]
RP   VARIANT THYROID TUMOR THR-592.
RX   MEDLINE; 97138061.
RA   SOARES P., BERX G., VAN ROY F., SOBRINHO-SIMOES M.;
RT   "E-cadherin gene alterations are rare events in thyroid tumors.";
RL   Int. J. Cancer 70:32-38(1997).
RN   [13]
RP   VARIANTS ASP-336 AND ILE-470.
RX   MEDLINE; 98196671.
RA   GUILFORD P., HOPKINS J., HARRAWAY J., MCLEOD M., MCLEOD N.,
RA   HARAWIRA P., TAITE H., SCOULAR R., MILLER A., REEVE A.E.;
RT   "E-cadherin germline mutations in familial gastric cancer.";
RL   Nature 392:402-405(1998).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT
CC       INVASIVE SUPPRESSOR ROLE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
CC   -!- DISEASE: DEFECTS OF CDH1 IS REGARDED AS ONE OF THE MAIN MOLECULAR
CC       EVENTS INVOLVED IN DYSFUNCTION OF THE CELL-CELL ADHESION SYSTEM,
CC       TRIGGERING CANCER INVASION (GASTRIC, BREAST, OVARY, ENDOMETRIUM
CC       AND THYROID) AND METASTASIS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z13009; CAA78353.1; -.
DR   EMBL; Z18923; CAA79356.1; -.
DR   EMBL; X12790; CAA31279.1; -.
DR   EMBL; L34545; AAA21764.1; -.
DR   PIR; S25141; IJHUCE.
DR   PIR; S37654; S37654.
DR   HSSP; P09803; 1SUH.
DR   MIM; 192090; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal; Disease mutation; Polymorphism.
FT   SIGNAL       1    27       POTENTIAL.
FT   PROPEP      28   154       POTENTIAL.
FT   CHAIN      155   882       EPITHELIAL-CADHERIN.
FT   DOMAIN     155   707       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   708   731       POTENTIAL.
FT   DOMAIN     732   882       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     155   262       CADHERIN 1.
FT   REPEAT     263   375       CADHERIN 2.
FT   REPEAT     376   486       CADHERIN 3.
FT   REPEAT     487   593       CADHERIN 4.
FT   REPEAT     594   697       CADHERIN 5.
FT   DOMAIN     838   851       SER-RICH.
FT   CARBOHYD   558   558       POTENTIAL.
FT   CARBOHYD   637   637       POTENTIAL.
FT   VARIANT    123   123       H -> Y (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001306.
FT   VARIANT    193   193       T -> P (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001307.
FT   VARIANT    274   277       MISSING (IN GASTRIC ADENOCARCINOMA).
FT                              /FTId=VAR_001308.
FT   VARIANT    315   315       N -> S (IN LOBULAR BREAST CARCINOMA).
FT                              /FTId=VAR_001309.
FT   VARIANT    336   336       E -> D.
FT                              /FTId=VAR_001310.
FT   VARIANT    370   370       D -> A (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001311.
FT   VARIANT    400   400       MISSING (IN GASTRIC CARCINOMA; LOSS OF
FT                              HETEROZYGOSITY).
FT                              /FTId=VAR_001312.
FT   VARIANT    418   423       MISSING (IN GASTRIC CARCINOMA).
FT                              /FTId=VAR_001313.
FT   VARIANT    463   463       E -> Q (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001314.
FT   VARIANT    470   470       T -> I.
FT                              /FTId=VAR_001315.
FT   VARIANT    470   470       T -> L (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001316.
FT   VARIANT    473   473       V -> N (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001317.
FT   VARIANT    592   592       A -> T (IN THYROID CANCER).
FT                              /FTId=VAR_001318.
FT   VARIANT    598   598       R -> Q (IN DIFFUSE GASTRIC CANCER).
FT                              /FTId=VAR_001319.
FT   VARIANT    617   617       A -> T (IN ENDOMETRIAL CANCER; LOSS OF
FT                              HETEROZYGOSITY).
FT                              /FTId=VAR_001320.
FT   VARIANT    711   711       L -> V (IN ENDOMETRIAL CANCER).
FT                              /FTId=VAR_001321.
FT   VARIANT    838   838       S -> G (IN OVARIAN CANCER; LOSS OF
FT                              HETEROZYGOSITY).
FT                              /FTId=VAR_001322.
FT   CONFLICT   543   543       S -> F (IN REF. 2).
SQ   SEQUENCE   882 AA;  97456 MW;  10821AB9 CRC32;
```

```
  Query Match           54.9%;  Score 39;  DB 1;  Length 882;
  Best Local Similarity  53.8%;  Pred. No. 13;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |: ||: |:||:|
Db     297 NAAIAYTILSQDP 309
```

```
RESULT   9
CAD2_HUMAN
ID   CAD2_HUMAN     STANDARD;      PRT;   906 AA.
AC   P19022; Q14923;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NEURAL-CADHERIN PRECURSOR (N-CADHERIN).
GN   CDH2 OR CDHN OR NCAD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91016946.
RA   REID R.A., HEMPERLY J.J.;
RT   "Human N-cadherin: nucleotide and deduced amino acid sequence.";
RL   Nucleic Acids Res. 18:5896-5896(1990).
RN   [2]
RP   REVISIONS TO 341; 699 AND 705.
RA   REID R.A.;
```

```
RL   Submitted (NOV-1990) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92363956.
RA   SALOMON D., AYALON O., PATEL-KING R., HYNES R.O., GEIGER B.;
RT   "Extrajunctional distribution of N-cadherin in cultured human
RT   endothelial cells.";
RL   J. Cell Sci. 102:7-17(1992).
RN   [4]
RP   SEQUENCE OF 160-906 FROM N.A.
RX   MEDLINE; 90347462.
RA   WALSH F.S., BARTON C.H., PUTT W., MOORE S.E., KELSELL D.,
RA   SPURR N., GOODFELLOW P.N.;
RT   "N-cadherin gene maps to human chromosome 18 and is not linked to the
RT   E-cadherin gene.";
RL   J. Neurochem. 55:805-812(1990).
RN   [5]
RP   SEQUENCE OF 1-20 FROM N.A.
RX   MEDLINE; 95048366.
RA   WALLIS J.A., FOX M., WALSH F.S.;
RT   "Structure of the human N-cadherin gene: YAC analysis and fine
RT   chromosomal mapping to 18q11.2.";
RL   Genomics 22:172-179(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57548; CAA40773.1; -.
DR   EMBL; X54315; CAA38213.1; -.
DR   EMBL; S42303; AAB22854.1; -.
DR   EMBL; M34064; AAA03236.1; -.
DR   EMBL; Z27420; CAA81799.1; -.
DR   PIR; A38870; IJHUCN.
DR   HSSP; P15116; 1NCH.
DR   MIM; 114020; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24    159
FT   CHAIN       160   906      NEURAL-CADHERIN.
FT   DOMAIN      160   724      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    725   746      POTENTIAL.
FT   DOMAIN      747   906      CYTOPLASMIC (POTENTIAL).
FT   REPEAT      160   267      CADHERIN 1.
FT   REPEAT      268   382      CADHERIN 2.
FT   REPEAT      383   497      CADHERIN 3.
FT   REPEAT      498   603      CADHERIN 4.
FT   REPEAT      604   714      CADHERIN 5.
FT   DOMAIN      863   878      SER-RICH.
FT   CARBOHYD    190   190      POTENTIAL.
FT   CARBOHYD    273   273      POTENTIAL.
FT   CARBOHYD    325   325      POTENTIAL.
FT   CARBOHYD    402   402      POTENTIAL.
FT   CARBOHYD    572   572      POTENTIAL.
FT   CARBOHYD    651   651      POTENTIAL.
FT   CARBOHYD    692   692      POTENTIAL.
FT   CONFLICT     12    12      MISSING (IN REF. 3).
FT   CONFLICT     16    16      L -> A (IN REF. 3 AND 5).
FT   CONFLICT    196   196      S -> T (IN REF. 1).
FT   CONFLICT    212   212      I -> L (IN REF. 4).
FT   CONFLICT    357   357      N -> I (IN REF. 1).
FT   CONFLICT    867   867      A -> L (IN REF. 3).
SQ   SEQUENCE   906 AA;  99851 MW;  BB1F9558 CRC32;


  Query Match             54.9%;  Score 39;  DB 1;  Length 906;
  Best Local Similarity   46.7%;  Pred. No. 14;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEPA 15
           ||  : ::|||| |:
Db     300 LNGMLRYRIVSQAPS 314


RESULT  10
VACB_MYCPN
ID   VACB_MYCPN     STANDARD;      PRT;   726 AA.
AC   P75529;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   VACB PROTEIN HOMOLOG.
GN   VACB.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA   HERRMANN R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000057; AAB96237.1; -.
DR   PFAM; PF00575; S1; 1.
DR   PFAM; PF00773; RNB; 1.
DR   PROSITE; PS01175; RIBONUCLEASE_II; 1.
KW   Hydrolase; Nuclease.
SQ   SEQUENCE   726 AA;  83219 MW;  C8974499 CRC32;


  Query Match             54.9%;  Score 39;  DB 1;  Length 726;
  Best Local Similarity   50.0%;  Pred. No. 11;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEPA 15
           |:|:||:|   :||
Db     193 NTKVAFQIDEYDPA 206


RESULT  11
CADB_XENLA
ID   CADB_XENLA     STANDARD;      PRT;   905 AA.
AC   P33152;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   BLASTOMERE-CADHERIN PRECURSOR (B-CADHERIN).
OS   Xenopus laevis (African clawed frog).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95151580.
RA   MULLER H.A., KUHL M., FINNEMANN S., SCHNEIDER S., VAN DER POEL S.,
RA   HAUSEN P., WEDLICH D.;
RT   "Xenopus cadherins: the maternal pool comprises distinguishable
RT   members of the family.";
RL   Mech. Dev. 47:213-223(1994).
RN   [2]
RP   SEQUENCE OF 459-905 FROM N.A.
RX   MEDLINE; 92062581.
RA   HERZBERG F., WILDERMUTH V., WEDLICH D.;
RT   "Expression of XBcad, a novel cadherin, during oogenesis and early
RT   development of Xenopus.";
RL   Mech. Dev. 35:33-42(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PITUITARY GLAND, LUNG AND KIDNEY.
CC   -!- DEVELOPMENTAL STAGE: DURING OOGENESIS AND EARLY DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X78546; CAA55292.1; -.
DR   EMBL; X63719; CAA45251.1; -.
DR   PIR; S43065; S43065.
DR   HSSP; P09803; 1SUH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     ?       POTENTIAL.
FT   PROPEP        ?   178       POTENTIAL.
FT   CHAIN       179   905       BLASTOMERE-CADHERIN.
FT   DOMAIN      179   727       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    728   751       POTENTIAL.
FT   DOMAIN      752   905       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      849   873       SER-RICH.
FT   DOMAIN      900   905       ASP/GLU-RICH (ACIDIC).
FT   CARBOHYD    448   448       POTENTIAL.
FT   CARBOHYD    581   581       POTENTIAL.
FT   CARBOHYD    704   704       POTENTIAL.
FT   CONFLICT    459   461       ILT -> NSA (IN REF. 2).
FT   CONFLICT    698   698       R -> Q (IN REF. 2).
FT   CONFLICT    808   808       V -> A (IN REF. 2).
FT   CONFLICT    841   841       D -> N (IN REF. 2).
FT   CONFLICT    878   878       D -> N (IN REF. 2).
FT   CONFLICT    884   884       D -> N (IN REF. 2).
FT   CONFLICT    903   903       MISSING (IN REF. 2).
SQ   SEQUENCE   905 AA;  100377 MW;  147E1228 CRC32;
```

```
  Query Match             53.5%;  Score 38;  DB 1;  Length 905;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
           ||  ||: |: |:|
Db     320 LNGVIAYSILKQDP 333
```

```
RESULT  12
CADN_XENLA
ID   CADN_XENLA     STANDARD;      PRT;   905 AA.
AC   P20310;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAL-CADHERIN 1 PRECURSOR (N-CADHERIN 1).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90211966.
RA   DETRICK R.J., DICKEY D., KINTNER C.R.;
RT   "The effects of N-cadherin misexpression on morphogenesis in Xenopus
RT   embryos.";
RL   Neuron 4:493-506(1990).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR   PIR; JQ0442; IJXLC1.
DR   HSSP; P15116; 1NCH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   PROPEP       29    160       POTENTIAL.
FT   CHAIN       161    905       NEURAL-CADHERIN 1.
FT   DOMAIN      161    723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    724    745       POTENTIAL.
FT   DOMAIN      746    905       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      161    268       CADHERIN 1.
FT   REPEAT      269    383       CADHERIN 2.
FT   REPEAT      384    498       CADHERIN 3.
FT   REPEAT      499    604       CADHERIN 4.
FT   REPEAT      605    713       CADHERIN 5.
FT   DOMAIN      862    877       SER-RICH.
FT   CARBOHYD    191    191       POTENTIAL.
FT   CARBOHYD    274    274       POTENTIAL.
FT   CARBOHYD    326    326       POTENTIAL.
FT   CARBOHYD    403    403       POTENTIAL.
FT   CARBOHYD    573    573       POTENTIAL.
FT   CARBOHYD    623    623       POTENTIAL.
FT   CARBOHYD    651    651       POTENTIAL.
FT   CARBOHYD    692    692       POTENTIAL.
SQ   SEQUENCE   905 AA;  100549 MW;  50053CA0 CRC32;
```

```
  Query Match             53.5%;  Score 38;  DB 1;  Length 905;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEPA 15
           |  : :||:|| ||
Db     302 NGMLRYKILSQTPA 315
```

```
RESULT  13
CADO_XENLA
ID   CADO_XENLA     STANDARD;      PRT;   906 AA.
AC   P33147;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
```

```
DE   NEURAL-CADHERIN 2 PRECURSOR (N-CADHERIN 2).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91372132.
RA   GINSBERG D., DESIMONE D., GEIGER B.;
RT   "Expression of a novel cadherin (EP-cadherin) in unfertilized eggs
RT   and early Xenopus embryos.";
RL   Development 111:315-325(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57675; CAA40867.1; -.
DR   PIR; A43785; IJXLC2.
DR   HSSP; P15116; 1NCH.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL       1    28       POTENTIAL.
FT   PROPEP      29   160       POTENTIAL.
FT   CHAIN      161   906       NEURAL-CADHERIN 2.
FT   DOMAIN     161   724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   725   746       POTENTIAL.
FT   DOMAIN     747   906       CYTOPLASMIC (POTENTIAL).
FT   REPEAT     161   268       CADHERIN 1.
FT   REPEAT     269   383       CADHERIN 2.
FT   REPEAT     384   498       CADHERIN 3.
FT   REPEAT     499   604       CADHERIN 4.
FT   REPEAT     605   714       CADHERIN 5.
FT   DOMAIN     863   878       SER-RICH.
FT   CARBOHYD   274   274       POTENTIAL.
FT   CARBOHYD   326   326       POTENTIAL.
FT   CARBOHYD   403   403       POTENTIAL.
FT   CARBOHYD   573   573       POTENTIAL.
FT   CARBOHYD   623   623       POTENTIAL.
FT   CARBOHYD   652   652       POTENTIAL.
FT   CARBOHYD   693   693       POTENTIAL.
SQ   SEQUENCE   906 AA;  100392 MW;  F0EB618C CRC32;


  Query Match             53.5%;  Score 38;  DB 1;  Length 906;
  Best Local Similarity   50.0%;  Pred. No. 21;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEPA 15
           |  : :||:|| ||
Db     302 NGMLRYKILSQTPA 315


RESULT  14
YOX1_YEAST
ID   YOX1_YEAST     STANDARD;      PRT;   385 AA.
AC   P34161;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HOMEOBOX PROTEIN YOX1.
GN   YOX1 OR YML027W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE OF 1-296 FROM N.A.
RX   MEDLINE; 93209080.
RA   KAUFMANN E.;
RT   "In vitro binding to the leucine tRNA gene identifies a novel yeast
RT   homeobox gene.";
RL   Chromosoma 102:174-179(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IN VITRO, IS CAPABLE OF BINDING TO THE DNA OF THE
CC       LEUCINE TRNA GENE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62392; CAA44264.1; -.
DR   EMBL; Z46659; CAA86628.1; -.
DR   PIR; S33388; S33388.
DR   HSSP; P06601; 1FJL.
DR   SGD; L0002540; YOX1.
DR   PFAM; PF00046; homeobox; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein.
FT   DNA_BIND    176   235       HOMEOBOX.
FT   CONFLICT    291   296       PRPHHS -> QGLIIP (IN REF. 1).
SQ   SEQUENCE   385 AA;  42739 MW;  1E5EB6A9 CRC32;


  Query Match             53.5%;  Score 38;  DB 1;  Length 385;
  Best Local Similarity   63.6%;  Pred. No. 8.6;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 KIAFKIVSQEP 14
           |  |||| |:|
Db     364 KFGFKIVDQQP 374


RESULT  15
DHAG_HUMAN
ID   DHAG_HUMAN     STANDARD;      PRT;   493 AA.
AC   P49189;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ALDEHYDE DEHYDROGENASE, E3 ISOZYME (EC 1.2.1.3) (GAMMA-
DE   AMINOBUTYRALDEHYDE DEHYDROGENASE) (EC 1.2.1.19) (R-AMINOBUTYRALDEHYDE
DE   DEHYDROGENASE).
GN   ALDH9 OR ALDH7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 96374830.
RA   LIN S.W., CHEN J.C., HSU L.C., HSIEH C.-L., YOSHIDA A.;
```

```
RT   "Human gamma-aminobutyraldehyde dehydrogenase (ALDH9): cDNA sequence,
RT   genomic organization, polymorphism, chromosomal localization, and
RT   tissue expression.";
RL   Genomics 34:376-380(1996).
RN   [2]
RP   SEQUENCE OF 32-493 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 94094820.
RA   KURYS G., SHAH P.C., KIKONYOGO A., REED D., AMBROZIAK W.,
RA   PIETRUSZKO R.;
RT   "Human aldehyde dehydrogenase. cDNA cloning and primary structure of
RT   the enzyme that catalyzes dehydrogenation of 4-aminobutyraldehyde.";
RL   Eur. J. Biochem. 218:311-320(1993).
RN   [3]
RP   CHARACTERIZATION.
RC   TISSUE=LIVER;
RX   MEDLINE; 89174735.
RA   KURYS G., AMBROZIAK W., PIETRUSZKO R.;
RT   "Human aldehyde dehydrogenase. Purification and characterization of a
RT   third isozyme with low Km for gamma-aminobutyraldehyde.";
RL   J. Biol. Chem. 264:4715-4721(1989).
CC   -!- FUNCTION: HAS A WIDE SUBSTRATE SPECIFICITY AND A LOW KM FOR 4-
CC       AMINOBUTYRALDEHYDE. ITS OPTIMAL PH DEPENDS UPON SUBSTRATE.
CC   -!- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H(2)O = ACID + NADH.
CC   -!- CATALYTIC ACTIVITY: 4-AMINOBUTANAL + NAD(+) + H(2)O = 4-
CC       AMINOBUTANOATE + NADH.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- TISSUE SPECIFICITY: HIGH EXPRESSION IN ADULT LIVER, SKELETAL
CC       MUSCLE, AND KIDNEY. LOW LEVELS IN HEART, PANCREAS, LUNG, AND
CC       BRAIN.
CC   -!- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN HUMAN EMBRYONIC BRAIN
CC       (GESTATIONAL AGE 12 WEEKS).
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34252; AAB18827.1; -.
DR   EMBL; X75425; CAA53176.1; -.
DR   HSSP; P56533; 1A4S.
DR   MIM; 602733; -.
DR   PFAM; PF00171; aldedh; 1.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW   Oxidoreductase; NAD.
FT   NP_BIND     231    236       NAD (ADP PART) (BY SIMILARITY).
FT   ACT_SITE    253    253       POTENTIAL.
FT   ACT_SITE    287    287       POTENTIAL.
FT   CONFLICT    115    115       C -> S (IN REF. 2).
FT   CONFLICT    149    149       C -> Q (IN AA SEQUENCE).
FT   CONFLICT    158    158       P -> W (IN AA SEQUENCE).
FT   CONFLICT    171    171       A -> R (IN REF. 2).
SQ   SEQUENCE   493 AA;  53532 MW;  97C6E858 CRC32;


  Query Match             52.1%;  Score 37;  DB 1;  Length 493;
  Best Local Similarity   66.7%;  Pred. No. 17;
  Matches    8;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQE 13
           |:| |||| ||:
Db      56 NAKAAFKIWSQK 67


Search completed: January  7, 2000, 13:25:38
Job time: 390 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:19 ; Search time 59.21 Seconds
(without alignments)
15.518 Million cell updates/sec

Title: US-08-991-628-3
Perfect score: 71
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database : SPTREMBL_11:*

Word size : 0

Number of hits that pass the threshold : 199794
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	97.2	993	11	O35902	O35902 mus musculu
2	40	56.3	327	5	Q93345	Q93345 caenorhabdi
3	40	56.3	145	6	O18926	O18926 sus scrofa
4	40	56.3	1421	10	O80907	O80907 arabidopsis
5	39	54.9	829	4	O75309	O75309 homo sapien
6	39	54.9	878	4	Q15855	Q15855 homo sapien
7	39	54.9	829	6	Q28634	Q28634 oryctolagus
8	38	53.5	533	2	Q54891	Q54891 streptococc
9	38	53.5	533	2	Q54989	Q54989 streptococc
10	38	53.5	138	2	P72670	P72670 synechocyst
11	38	53.5	967	4	O94853	O94853 homo sapien
12	38	53.5	273	9	Q38172	Q38172 bacteriopha
13	37	52.1	236	2	O05946	O05946 rickettsia
14	37	52.1	461	2	Q9ZCY0	Q9zcy0 rickettsia
15	37	52.1	936	4	O95206	O95206 homo sapien
16	37	52.1	252	12	Q9YW59	Q9yw59 melanoplus
17	36	50.7	103	2	O53450	O53450 mycobacteri
18	36	50.7	475	2	Q9ZCY9	Q9zcy9 rickettsia
19	36	50.7	238	11	O55075	O55075 cricetulus
20	36	50.7	830	11	O88338	O88338 mus musculu
21	36	50.7	277	11	O08558	O08558 rattus norv
22	36	50.7	276	11	O08559	O08559 rattus norv
23	36	50.7	906	11	Q9Z1Y3	Q9z1y3 rattus norv
24	35	49.3	308	2	O51057	O51057 borrelia bu
25	35	49.3	182	5	Q25918	Q25918 plasmodium
26	35	49.3	225	5	Q20152	Q20152 caenorhabdi
27	35	49.3	241	8	Q33766	Q33766 allomyces m
28	35	49.3	838	9	O64330	O64330 bacteriopha
29	35	49.3	1249	10	O23289	O23289 arabidopsis
30	35	49.3	2261	10	Q42793	Q42793 glycine max
31	35	49.3	775	10	P93205	P93205 lycopersico
32	35	49.3	2228	12	O55528	O55528 sendai viru
33	35	49.3	2228	12	O55530	O55530 sendai viru
34	35	49.3	2842	12	O36452	O36452 hepatitis g
35	35	49.3	1980	12	Q84185	Q84185 human parai
36	35	49.3	2228	12	Q98705	Q98705 sendai viru
37	34	47.9	334	1	O58135	O58135 pyrococcus
38	34	47.9	380	2	O66815	O66815 aquifex aeo
39	34	47.9	330	2	O87491	O87491 staphylococ
40	34	47.9	455	3	O14039	O14039 schizosacch
41	34	47.9	1272	3	O13756	O13756 schizosacch
42	34	47.9	464	4	O43159	O43159 homo sapien
43	34	47.9	591	4	O60324	O60324 homo sapien
44	34	47.9	716	4	Q14164	Q14164 homo sapien
45	34	47.9	813	4	O75229	O75229 homo sapien

ALIGNMENTS

RESULT 1
O35902
ID O35902 PRELIMINARY; PRT; 993 AA.
AC O35902;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DESMOGLEIN 3 (FRAGMENT).
GN DSG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ISHIKAWA H., LI K., UITTO J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; U86016; AAB65091.1; -.
DR PFAM; PF00028; cadherin; 4.
DR PROSITE; PS00232; CADHERIN; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 97.2%; Score 69; DB 11; Length 993;
Best Local Similarity 93.3%; Pred. No. 9.9e-05;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LNSKIAFKIVSQEPA 15
:|||||||||||||
Db 190 MNSKIAFKIVSQEPA 204

RESULT 2
Q93345
ID Q93345 PRELIMINARY; PRT; 327 AA.
AC Q93345;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE C36B1.11 PROTEIN.
GN C36B1.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]

```
RP   SEQUENCE FROM N.A.
RA   LENNARD N.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z80215; CAB02276.1; -.
SQ   SEQUENCE   327 AA;  37139 MW;  D8A2C259 CRC32;


  Query Match             56.3%;  Score 40;  DB 5;  Length 327;
  Best Local Similarity   80.0%;  Pred. No. 9.1;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 KIAFKIVSQE 13
           |:|||| |||
Db      81 KLAFKIASQE 90


RESULT   3
O18926
ID   O18926      PRELIMINARY;      PRT;   145 AA.
AC   O18926;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   P-CADHERIN (FRAGMENT).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LUTZ D.A., ZHENG J.J.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF033826; AAB87087.1; -.
DR   PFAM; PF00028; cadherin; 1.
FT   NON_TER      1      1
FT   NON_TER    145    145
SQ   SEQUENCE   145 AA;  15617 MW;  49B7A084 CRC32;


  Query Match             56.3%;  Score 40;  DB 6;  Length 145;
  Best Local Similarity   53.8%;  Pred. No. 3.9;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |  :|: |:||||
Db      74 NGVVAYSILSQEP 86


RESULT   4
O80907
ID   O80907      PRELIMINARY;      PRT;  1421 AA.
AC   O80907;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   T19C21.7 PROTEIN.
GN   T19C21.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004683; AAC28760.1; -.
SQ   SEQUENCE   1421 AA;  154325 MW;  1B6D2FF1 CRC32;


  Query Match             56.3%;  Score 40;  DB 10;  Length 1421;
  Best Local Similarity   42.9%;  Pred. No. 42;
  Matches    6;  Conservative    7;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQEP 14
             ::|||:::|::| |
Db     262 VDSKISYEIITQNP 275


RESULT   5
O75309
ID   O75309      PRELIMINARY;      PRT;   829 AA.
AC   O75309;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KSP-CADHERIN.
GN   CDH16.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98389630.
RA   THOMSON R.B., WARD D.C., QUAGGIN S.E., IGARASHI P., MUCKLER Z.E.,
RA   ARONSON P.S.;
RT   "cDNA cloning and chromosomal localization of the human and mouse
RT   isoforms of Ksp-cadherin.";
RL   Genomics 51:445-451(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF016272; AAC34255.1; -.
DR   PFAM; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN; 2.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   829 AA;  89923 MW;  E44B3DBE CRC32;


  Query Match             54.9%;  Score 39;  DB 4;  Length 829;
  Best Local Similarity   57.1%;  Pred. No. 37;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEPA 15
           || : | |:|| ||
Db     160 NSDLRFHILSQAPA 173


RESULT   6
Q15855
ID   Q15855      PRELIMINARY;      PRT;   878 AA.
AC   Q15855; Q16194; Q13799;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   UVOMORULIN PRECURSOR (E-CADHERIN) (ARC-1/UVOMORULIN).
GN   UVO.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 94242050.
RA   RIMM D.L., MORROW J.S.;
RT   "Molecular cloning of human E-cadherin suggests a novel subdivision
RT   of the cadherin superfamily.";
RL   Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
RN   [2]
RP   SEQUENCE OF 333-472 FROM N.A.
RX   MEDLINE; 94306394.
RA   BECKER K.F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
RA   SIEWERT J.R., HOFLER H.;
RT   "E-cadherin gene mutations provide clues to diffuse type gastric
RT   carcinomas.";
RL   Cancer Res. 54:3845-3852(1994).
RN   [3]
RP   SEQUENCE OF 261-388 FROM N.A.
RC   TISSUE=LIVER;
RA   FRIXEN U.H.;
RL   Submitted (MAR-1990) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; S72492; AAD14108.1; -.
DR   EMBL; S72397; AAD14108.1; JOINED.
DR   EMBL; S72491; AAD14108.1; JOINED.
DR   EMBL; L08599; AAA61259.1; -.
DR   EMBL; X52279; CAA36522.1; -.
DR   PFAM; PF00028; cadherin; 5.
DR   PFAM; PF01049; Cadherin_C_term; 1.
DR   PROSITE; PS00232; CADHERIN; 3.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    878       UVOMORULIN.
SQ   SEQUENCE   878 AA;  96741 MW;  8F7F0180 CRC32;

  Query Match             54.9%;  Score 39;  DB 4;  Length 878;
  Best Local Similarity   53.8%;  Pred. No. 39;
  Matches    7;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEP 14
           |: ||: |:||:|
Db     293 NAAIAYTILSQDP 305


RESULT   7
Q28634
ID   Q28634      PRELIMINARY;      PRT;   829 AA.
AC   Q28634;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   KIDNEY-SPECIFIC CADHERIN PRECURSOR (KSP-CADHERIN).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A., AND OF SEQUENCE 570-586.
RC   STRAIN=NEW ZEALAND WHITE;
RX   MEDLINE; 95340560.
RA   THOMSON R.B., IGARASHI P., BIEMESDERFER D., KIM R., ABU-ALFA A.,
RA   SOLEIMANI M., ARONSON P.S.;
RT   "Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific
RT   member of the cadherin multigene family.";
RL   J. Biol. Chem. 270:17594-17601(1995).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: KIDNEY.
CC       LIMITED TO THE BASOLATERAL MEMBRANES OF RENAL TUBULAR EPITHELIAL
CC       CELLS.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC       MOST CLOSELY RESEMBLES MEMBERS OF THE LI-CADHERIN/HPT-1 CADHERIN
CC       SUBGROUP.
DR   EMBL; U28945; AAC48472.1; -.
DR   PROSITE; PS00232; CADHERIN; 2.
DR   PFAM; PF00028; cadherin; 6.
KW   Cell adhesion; Multigene family; Glycoprotein; Repeat; Signal;
KW   Calcium-binding; Transmembrane.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    829       KIDNEY-SPECIFIC CADHERIN.
FT   DOMAIN       19    786       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    787    807       POTENTIAL.
FT   DOMAIN      808    829       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       18    124       ECTODOMAIN A.
FT   REPEAT      125    240       ECTODOMAIN B.
FT   REPEAT      241    334       CADHERIN 1 (ECTODOMAIN C).
FT   REPEAT      335    447       CADHERIN 2 (ECTODOMAIN D).
FT   REPEAT      448    562       CADHERIN 3 (ECTODOMAIN E).
FT   REPEAT      563    664       CADHERIN 4 (ECTODOMAIN F).
FT   REPEAT      665    786       CADHERIN 5 (ECTODOMAIN G).
FT   CARBOHYD    517    517       POTENTIAL.
FT   CARBOHYD    602    602       POTENTIAL.
FT   CARBOHYD    709    709       POTENTIAL.
FT   CARBOHYD    722    722       POTENTIAL.
SQ   SEQUENCE   829 AA;  88828 MW;  7C163641 CRC32;

  Query Match             54.9%;  Score 39;  DB 6;  Length 829;
  Best Local Similarity   57.1%;  Pred. No. 37;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 NSKIAFKIVSQEPA 15
           || : | |:|| ||
Db     160 NSDLRFHILSQTPA 173


RESULT   8
Q54891
ID   Q54891      PRELIMINARY;      PRT;   533 AA.
AC   Q54891;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MGA4 POSITIVE REGULATORY PROTEIN.
GN   MGA4.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AP4;
RX   MEDLINE; 97074652.
RA   ANDERSSON G., MCIVER K., HEDEN L.O., SCOTT J.R.;
RT   "Complementation of divergent mga genes in group A Streptococcus.";
RL   Gene 175:77-81(1996).
DR   EMBL; X95408; CAA64690.1; -.
SQ   SEQUENCE   533 AA;  62099 MW;  BFA145A3 CRC32;

  Query Match             53.5%;  Score 38;  DB 2;  Length 533;
  Best Local Similarity   44.4%;  Pred. No. 36;
  Matches    8;  Conservative    5;  Mismatches    1;  Indels    4;  Gaps    1;

Qy       1 LNSKI----AFKIVSQEP 14
           || |:    ||:|::|:|
Db     271 LNDKLEIGCAFEIINQDP 288


RESULT   9
Q54989
ID   Q54989      PRELIMINARY;      PRT;   533 AA.
AC   Q54989;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   VIRR GENE.
GN   VIRR49.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE M49;
RA   PODBIELSKI A.;
RL   Submitted (SEP-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-499 FROM N.A.
RC   STRAIN=SEROTYPE M49;
RX   MEDLINE; 95105032.
RA   PODBIELSKI A., FLOSDORFF A., WEBERHEYNEMANN J.;
RT   "The group A streptococcal virR49 gene controls expression of four
RT   structural vir regulon genes.";
RL   Infect. Immun. 63:9-20(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE M49;
RA   PODBIELSKI A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X68501; CAA48513.1; -.
SQ   SEQUENCE   533 AA;  62164 MW;  F46E7EF2 CRC32;

  Query Match             53.5%;  Score 38;  DB 2;  Length 533;
  Best Local Similarity   44.4%;  Pred. No. 36;
  Matches    8;  Conservative    5;  Mismatches    1;  Indels    4;  Gaps    1;

Qy       1 LNSKI----AFKIVSQEP 14
           || |:    ||:|::|:|
Db     271 LNDKLEIGCAFEIINQDP 288


RESULT  10
P72670
ID   P72670      PRELIMINARY;      PRT;   138 AA.
AC   P72670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 16.6 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90899; BAA16672.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   138 AA;  16580 MW;  CA436352 CRC32;

  Query Match             53.5%;  Score 38;  DB 2;  Length 138;
  Best Local Similarity   41.7%;  Pred. No. 8.9;
  Matches    5;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SKIAFKIVSQEP 14
           :|:||::: |:|
Db      28 AKLAPEVIPQDP 39


RESULT  11
O94853
ID   O94853      PRELIMINARY;      PRT;   967 AA.
AC   O94853;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   KIAA0753 PROTEIN.
GN   KIAA0753.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 99087487.
RA   NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., MIYAJIMA N., TANAKA A.,
RA   KOTANI H., NOMURA N., OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
DR   EMBL; AB018296; BAA34473.1; -.
SQ   SEQUENCE   967 AA;  109376 MW;  C8749C0A CRC32;

  Query Match             53.5%;  Score 38;  DB 4;  Length 967;
  Best Local Similarity   46.2%;  Pred. No. 67;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LNSKIAFKIVSQE 13
           | | ::| ::|||
Db      83 LGSSVSFSVISQE 95


RESULT  12
Q38172
ID   Q38172      PRELIMINARY;      PRT;   273 AA.
AC   Q38172;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MAJOR HEAD PROTEIN.
GN   MHP.
OS   Bacteriophage B1.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TP901-1, TP936-1, C3-T1;
RX   MEDLINE; 96193742.
RA   JOHNSEN M.G., APPEL K.F., MADSEN P.L., VOGENSEN F.K., HAMMER K.,
RA   ARNAU J.;
RT   "A genomic region of lactococcal temperate bacteriophage TP901-1
RT   encoding major virion proteins.";
RL   Virology 218:306-315(1996).
DR   EMBL; X84706; CAA59185.1; -.
SQ   SEQUENCE   273 AA;  28880 MW;  E7DA3E03 CRC32;

  Query Match             53.5%;  Score 38;  DB 9;  Length 273;
  Best Local Similarity   61.5%;  Pred. No. 18;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 SKIAFKIVSQEPA 15
           | : |||||  ||
```

```
Db    213 SALMFKIVSNSPA 225


RESULT  13
O05946
ID   O05946      PRELIMINARY;      PRT;   236 AA.
AC   O05946;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PHOSPHORIBOSYLAMINOIMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE.
GN   PURC OR RP220.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 97419517.
RA   ANDERSSON J.O., ANDERSSON S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE; 99039499.
RA   ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA   SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA   ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RA   ANDERSSON S.G.E.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y11780; CAA72468.1; -.
DR   EMBL; AJ235270; CAA14683.1; -.
DR   PFAM; PF01259; SAICAR_synt; 1.
SQ   SEQUENCE   236 AA;  27310 MW;  3123672C CRC32;


  Query Match             52.1%;  Score 37;  DB 2;  Length 236;
  Best Local Similarity   38.5%;  Pred. No. 24;
  Matches    5;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEP 14
          | |: |:::  ||
Db    207 NEKLGFELIQNEP 219


RESULT  14
Q9ZCY0
ID   Q9ZCY0      PRELIMINARY;      PRT;   461 AA.
AC   Q9ZCY0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SODIUM/PANTOTHENATE SYMPORTER (PANF).
GN   RP571.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 99039499.
RA   ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA   SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA   ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;
RT   "The Genome Sequence of Rickettsia prowazekii and the Origin of
RT   Mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RA   ANDERSSON S.G.E.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ235272; CAA15019.1; -.
SQ   SEQUENCE   461 AA;  51674 MW;  4D6121F9 CRC32;


  Query Match             52.1%;  Score 37;  DB 2;  Length 461;
  Best Local Similarity   46.7%;  Pred. No. 48;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 LNSKIAFKIVSQEPA 15
          || |||:   :|:
Db    276 LNGLIAFKLYPNQPS 290


RESULT  15
O95206
ID   O95206      PRELIMINARY;      PRT;   936 AA.
AC   O95206;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PROTOCADHERIN.
GN   PCDH8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99005535.
RA   STREHL S., GLATT K., LIU Q.M., GLATT H., LALANDE M.;
RT   "Characterization of two novel protocadherins (PCDH8 and PCDH9)
RT   localized on human chromosome 13 and mouse chromosome 14.";
RL   Genomics 53:81-89(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF061573; AAC70009.1; -.
DR   PROSITE; PS00232; CADHERIN; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   936 AA;  98879 MW;  3A6F1EC8 CRC32;


  Query Match             52.1%;  Score 37;  DB 4;  Length 936;
  Best Local Similarity   46.2%;  Pred. No. 99;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 NSKIAFKIVSQEP 14
          | ::|||::  |||
Db    646 NGELAFELQQQEP 658


Search completed: January 12, 2000, 23:15:21
Job time: 184 sec
```

```
CC   pemphigus vulgaris.
SQ   Sequence   614 AA;


  Query Match             100.0%;  Score 74;  DB 1;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          |||||||||||||||
Db     78 ATQKITYRISGVGID 92


RESULT   3
W04841
ID   W04841 standard; peptide; 15 AA.
AC   W04841;
DT   18-FEB-1997  (first entry)
DE   Self epitope of desmoglein 3, implicated in autoimmune disease.
KW   Tolerisation; self-epitope; antigen; autoimmune disease;
KW   autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW   pemphigus vulgaris; desmoglein; multiple sclerosis;
KW   herpes simplex virus; adenovirus; phosphomannomutase;
KW   human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW   influenza; haemagglutinin; reovirus; sigma protein.
OS   Homo sapiens.
PN   WO9627387-A1.
PD   12-SEP-1996.
PF   07-MAR-1996; U03182.
PR   07-MAR-1995; US-400796.
PA   (HARD ) HARVARD COLLEGE.
PI   Strominger JL,  Wucherpfennig KW;
DR   WPI; 96-425218/42.
PT   Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT   antigens - useful in disease treatment, and method for
PT   identification of other self and non-self antigens implicated in
PT   auto-immune disease
PS   Claim 1; Page 38; 58pp; English.
CC   Pharmaceutical preparations for tolerisation to antigens comprise
CC   either an isolated human non-collagen or non-myslin basic protein
CC   (MBP) polypeptide which is capable of tolerising an individual to an
CC   autoantigen; or an isolated human pathogen polypeptide capable of
CC   tolerising an individual to that polypeptide.  In both cases, the
CC   polypeptide (whether self or non-self) includes an amino acid
CC   sequence corresponding to a sequence motif for a MHC class II
CC   protein, such as HLA-DR, which is associated with a human autoimmune
CC   disease and which binds to the polypeptide to activate autoreactive
CC   T-cells in individuals with the autoimmune disease. This peptide is
CC   derived from the human desmoglein 3 protein (amino acids 78-93 (sic))
CC   and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC   derived from the human desmoglein protein are described in W04841-47.
SQ   Sequence   15 AA;


  Query Match             100.0%;  Score 74;  DB 1;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          |||||||||||||||
Db      1 ATQKITYRISGVGID 15


RESULT   4
W15489
ID   W15489 standard; Protein; 778 AA.
AC   W15489;
DT   17-JUN-1997  (first entry)
DE   Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW   Pemphigus foliaceus; autoantibody; constant region; IgG;
KW   extracellular region; antigen; hinge portion; skin;
KW   dermatitis herpetiformis; fusion protein; detection; ss.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          1. .545
FT                   /note= "Pemphigus foliaceus antigen protein"
PN   J09077800-A.
PD   25-MAR-1997.
PF   12-SEP-1995; 260899.
PR   12-SEP-1995; JP-260899.
PA   (NISH/) NISHIKAWA T.
DR   WPI; 97-241758/22.
DR   P-PSDB; T66428.
PT   Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT   through the hinge region used to treat pemphigus foliaceus
PS   Claim 1; Page 10-12; 17pp; Japanese.
CC   This sequence represents a fused protein recognised by pemphigus
CC   foliaceus patient autoantibody which comprises the constant region
CC   of IgG linked to the extracellular region of pemphigus foliaceus
CC   antigen protein through the hinge portion.  Pemphigus foliaceus is
CC   a chronic, generalised, vesicular and scaling skin eruption similar
CC   to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC   protein is useful to treat pemphigus foliaceus.  The antigen is
CC   especially administered through an adsorbent upon which the fusion
CC   protein is immobilised via a carrier. The fusion protein is also
CC   useful for detecting pemphigus foliaceus antibodies which is useful
CC   in immunodiagnosis.  The fusion protein has little or no side effects.
SQ   Sequence   778 AA;


  Query Match             86.5%;  Score 64;  DB 1;  Length 778;
  Best Local Similarity   80.0%;  Pred. No. 0.00073;
  Matches   12;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATQKITYRISGVGID 15
          | |::|||||||||||
Db     79 ANQQVTYRISGVGID 93


RESULT   5
R71376
ID   R71376 standard; Protein; 551 AA.
AC   R71376;
DT   21-NOV-1995  (first entry)
DE   Human cystathionine beta-synthase.
KW   Cystathionine; beta synthase; human; homocystinuria.
OS   Homo sapiens.
PN   WO9507714-A.
PD   23-MAR-1995.
PF   12-SEP-1994; U10203.
PR   13-SEP-1993; US-120960.
PA   (COLS ) UNIV COLORADO.
PI   Kraus JP;
DR   WPI; 95-131185/17.
DR   N-PSDB; Q87430.
PT   Purified DNA encoding human cystathionine beta-synthase - useful
PT   for producing human cystathionine beta-synthase, used for
PT   treating homocystinuria
PS   Claim 7; Fig 1a-c; 45pp; English.
CC   The amino acid sequence shown in R71376 is the human cystathionine
CC   beta-synthase (CBS). Human CBS can be used to treat patients with
CC   homocystinuria and the cDNA sequence, Q87430, from which it is
CC   derived is also useful for screening CBS deficient patients for
CC   mutations in the CBS gene.
SQ   Sequence   551 AA;


  Query Match             56.8%;  Score 42;  DB 1;  Length 551;
  Best Local Similarity   42.9%;  Pred. No. 6;
  Matches    6;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 TQKITYRISGVGID 15
          |:: || : |:|-|
Db    296 TEQTTYEVEGIGYD 309
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:12:33 ; Search time 51.33 Seconds
(without alignments)
6.922 Million cell updates/sec

Title: US-08-991-628-1
Perfect score: 74
Sequence: 1 ATQKITYRISGVGID 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	999	1	R30742	Human pemphigus vu
2	74	100.0	614	1	W07908	Pemphigus vulgaris
3	74	100.0	15	1	W04841	Self epitope of de
4	64	86.5	778	1	W15489	Pemphigus foliaceu
5	42	56.8	551	1	R71376	Human cystathionin
6	39	52.7	504	1	W60670	Human glaucoma ass
7	39	52.7	504	1	W64669	Human TIGR protein
8	39	52.7	504	1	W70496	Trabecular meshwor
9	39	52.7	504	1	W73500	Trabecular meshwor
10	39	52.7	504	1	W89391	Human trabecular m
11	37	50.0	309	1	R06441	MG-6 antigen. Diag
12	37	50.0	1363	1	R11061	Bovine Coronavirus
13	37	50.0	560	1	W13009	Segment of desmoso
14	37	50.0	1363	1	W31706	Bovine coronavirus
15	37	50.0	390	1	Y04860	Mycobacterium spec
16	36	48.6	712	1	R27823	Sequence encoded b
17	36	48.6	717	1	R27824	Sequence encoded b
18	36	48.6	1065	1	W33819	Arabidopsis cellul
19	36	48.6	289	1	Y06966	E. canis P30-6 pro
20	35	47.3	1560	1	P94145	S. cremoris protei
21	35	47.3	1962	1	R10560	Mutant protease (K
22	35	47.3	1962	1	R10561	Mutant protease (N
23	35	47.3	1962	1	R10557	Mutant protease (A
24	35	47.3	1962	1	R10558	Mutant protease (A
25	35	47.3	1962	1	R10559	Mutant protease (A
26	35	47.3	1959	1	R10562	Mutant protease (d
27	35	47.3	1962	1	R10563	Mutant protease (K
28	35	47.3	1974	1	R10940	Mutant protease (d
29	35	47.3	1968	1	R10941	Mutant protease (d
30	35	47.3	985	1	R42995	Glycosyltransferas
31	35	47.3	985	1	R42214	Aspergillus niger
32	35	47.3	878	1	R55060	Sequence of human
33	35	47.3	878	1	R85487	Human E-cadherin p
34	35	47.3	916	1	W13129	Full length human
35	35	47.3	916	1	W25658	Human cadherin-4.
36	35	47.3	626	1	W37140	Pea plastidial pho
37	34.5	46.6	593	1	R86292	Pseudomonas mesoac
38	34	45.9	304	1	P70637	Sequence encoded b
39	34	45.9	589	1	P60303	Sequence encoded b
40	34	45.9	896	1	R63533	Human HT-1376 cell
41	34	45.9	23	1	R65100	Random biotinylati
42	34	45.9	23	1	W46703	Biotinylation pept
43	34	45.9	626	1	W55115	Streptococcus pneu
44	33	44.6	100	1	R37660	Pseudomonas 7A glu
45	33	44.6	336	1	R59739	Pseudomonas glutam

ALIGNMENTS

RESULT 1
R30742
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 74; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATQKITYRISGVGID 15
||||||||||||||||
Db 79 ATQKITYRISGVGID 93

RESULT 2
W07908
ID W07908 standard; protein; 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphis vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphis vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucuous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose

```
CC   from the E2 initiation codon.  The E2 and E3 genes have been
CC   cloned and can be used for the recombinant production of BCV
CC   polypeptides, using e.g. Spodoptera frugiperda Sf9 insect cells
CC   as host cells.  Glycosylated and non-glycosylated recombinant
CC   E2 and E3 (see W31707) are useful as components of vaccines
CC   directed toward preventing BCV infection, or reducing the severity
CC   of BCV infection, in bovine populations.
SQ   Sequence   1363 AA;


  Query Match '           50.0%;  Score 37;  DB 1;  Length 1363;
  Best Local Similarity   50.0%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ITYRISGVGI 14
           : |||:|:|:
Db     996 VQYRINGIGV 1005


RESULT  15
Y04860
ID   Y04860 standard; Protein; 390 AA.
AC   Y04860;
DT   06-JUL-1999  (first entry)
DE   Mycobacterium species protein sequence 18F.
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
OS   Mycobacterium sp.
PN   WO9909186-A2.
PD   25-FEB-1999.
PF   14-AUG-1998; F01813.
PR   11-SEP-1997; FR-011325.
PR   14-AUG-1997; FR-010404.
PA   (INSP ) INST PASTEUR.
PI   Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y,
PI   Guigueno A;
DR   WPI; 99-181045/15.
DR   N-PSDB; X34112.
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
PS   Claim 32; Fig 18F; 309pp; French.
CC   Sequences Y04742-Y05000 and Y07201-Y07204 represent secreted proteins
CC   from various Mycobacterium species microorganisms.  The encoding
CC   nucleotide sequences can be used as primers and probes for methods
CC   for detecting and identifying mycobacteria, especially belonging to
CC   the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
SQ   Sequence   390 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 390;
  Best Local Similarity   50.0%;  Pred. No. 35;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 KITYRISGVGID 15
           ::| |: |:|||
Db      10 QVTGRVFGIGID 21


Search completed: January  6, 2000, 15:40:34
Job time: 1681 sec
```

```
PT   incorporated with its coding gene.
PS   Claim 2; Fig 1f; 20pp; Japanese.
CC   DNA encoding the protein can be inserted into an expression vector
CC   for the prodn. of MG-1 polypeptide which elicits an antigen-antibody
CC   reaction with anti-mycoplasma gallisepticum poultry sera.  It may also
CC   be ligated to other DNA to produce fusion proteins with an N-terminal
CC   bacterial enzyme sequence.
CC   See also R05081-2 and R06437-40.
SQ   Sequence   309 AA;

  Query Match            50.0%;  Score 37;  DB 1;  Length 309;
  Best Local Similarity  58.3%;  Pred. No. 27;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 TQKITYRISGVG 13
           || :| | :|||
Db     200 TQTVTNRFAGVG 211


RESULT  12
R11061
ID   R11061 standard; Protein; 1363 AA.
AC   R11061;
DT   24-MAY-1991  (first entry)
DE   Bovine Coronavirus E2 protein.
KW   Bovine Coronavirus; BCV; E2; E3; vaccine.
OS   Bovine coronavirus - Quebec isolate.
FH   Key             Location/Qualifiers
FT   peptide         1. .17
FT                   /label= sig_peptide
FT   cleavage_site   766. .766
FT                   /note= "probable cleavage site"
FT   domain          1306. .1338
FT                   /label= transmembrane_domain
PN   WO9102752-A.
PD   07-MAR-1991.
PF   10-AUG-1990; CA0252.
PR   22-AUG-1989; US-397689.
PA   (VETE-) VETERINARY INFECTIO.
PI   Parker MD, Cox GJ, Babiuk LA;
DR   WPI; 91-087247/12.
DR   N-PSDB; Q10947.
PT   Deoxyribonucleic acid encoding Bovine coronavirus protein E2 and
PT   E3 - useful as vaccine component
PS   Disclosure; Fig 3 (1-6); 70pp; English.
CC   E2 protein is one of the four proteins (N, E1, E2 and E3) composing
CC   BCV.  The E2 and E3 polypeptides, or fragments of these, are used as
CC   subunit antigens in vaccines for protection against or ameliorating
CC   BCV, without risk of infection.
SQ   Sequence   1363 AA;

  Query Match            50.0%;  Score 37;  DB 1;  Length 1363;
  Best Local Similarity  50.0%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ITYRISGVGI 14
           : |||:|:|:
Db     996 VQYRINGIGV 1005


RESULT  13
W13009
ID   W13009 standard; protein; 560 AA.
AC   W13009;
DT   21-NOV-1997  (first entry)
DE   Segment of desmosomal cadherin, desmoglein Dsg2.
KW   Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW   carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW   micrometastasis; separation; enrichment; targetted delivery;
KW   metastatic.
OS   Homo sapiens.
PN   DE19531033-A1.
PD   27-FEB-1997.
PF   23-AUG-1995; 031033.
PR   23-AUG-1995; DE-031033.
PA   (PROG-) PROGEN BIOTECHNIK GMBH.
PI   Franke WW,  Schaefer S;
DR   WPI; 97-146518/14.
PT   Antibody reactive with part of desmosomal cadherin - exposed on
PT   surface of epithelial or carcinoma cells, not bound to desmosomes,
PT   useful for diagnosis and treatment of carcinoma micrometastases
PS   Claim 7; Page 5; 8pp; German.
CC   The present sequence is a segment of the desmosomal cadherin (DC),
CC   desmoglein Dsg2, which is exposed on the surface of epithelial or
CC   carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC   directed against epitopes of the present sequence can be used to
CC   diagnose, i.e. to detect carcinoma cells, especially
CC   micrometastases, not bound to desmosomes, to separate, enrich or
CC   detect living or fixed carcinoma cells by cell sorting methods and
CC   as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC   target cells. The Ab provides rapid and reliable detection of
CC   metastatic carcinoma, and detects parts of DC that are not
CC   accessible in desmosome bound cells, as in normal tissue or
CC   carcinomas.
SQ   Sequence   560 AA;

  Query Match            50.0%;  Score 37;  DB 1;  Length 560;
  Best Local Similarity  63.6%;  Pred. No. 52;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 KITYRISGVGI 14
           ||||: :| ||
Db      35 KITYKYTGKGI 45


RESULT  14
W31706
ID   W31706 standard; Protein; 1363 AA.
AC   W31706;
DT   14-APR-1998  (first entry)
DE   Bovine coronavirus E2 (S) protein.
KW   BCV; E2 protein; peplomer protein; S spike; antigen; vaccine;
KW   cattle.
OS   Bovine coronavirus.
FH   Key             Location/Qualifiers
FT   Peptide         1. .17
FT                   /label= Sig_peptide
FT   Protein         18. .1363
FT                   /label= Mat_protein
FT   Domain          1306. .1338
FT                   /note= "transmembrane domain"
PN   US5672350-A.
PD   30-SEP-1997.
PF   22-AUG-1989; 397689.
PR   19-DEC-1991; US-811422.
PR   22-AUG-1989; US-397689.
PR   18-OCT-1991; US-779500.
PR   22-DEC-1993; US-171763.
PA   (VETE-) VETERINARY INFECTIOUS DISEASE.
PI   Babiuk LA, Cox GJ, Parker MD;
DR   WPI; 97-488823/45.
DR   N-PSDB; T89387.
PT   Vaccines against bovine coronavirus - containing recombinant bovine
PT   coronavirus polypeptide(s)
PS   Claim 13; Fig 3; 52pp; English.
CC   This polypeptide comprises the E2 protein, also designated
CC   peplomer protein or S (Spike), of bovine coronavirus (BCV).  It
CC   has a mol.wt. of 150 kDa exclusive of glycosylation and contains
CC   21 potential N-linked glycosylation sites.  The amino acid
CC   sequence was deduced from an clone E2 cDNA (see T89387).  The
CC   E2 gene in plasmid pT18E2 (E. coli JM105) is deposited as ATCC
CC   68041.  The BCV E3 gene (see T89388) is immediately 5' of the E2
CC   gene on the viral genome and terminates 14 nucleotides upstream
```

```
PA   (REGC ) UNIV CALIFORNIA.
PI   Huang W, Nguyen TD, Polansky JR;
DR   WPI; 98-542701/46.
DR   N-PSDB; V33484.
PT   New protein induced in trabecular meshwork cells by glucocorticoids
PT   - useful in the diagnosis of glaucoma and related diseases
PS   Claim 1; Fig 1A-C; 53pp; English.
CC   This is the amino acid sequence of a new human 55 kDa protein,
CC   designated trabecular meshwork induced glucocorticoid response*
CC   (TIGR*) protein, that is highly induced by glucocorticoids in the
CC   endothelial lining of the human trabecular meshwork (HTM). The
CC   sequence was deduced from an isolated cDNA clone (see V33484).
CC   Studies of the recombinant protein suggest (1) that the 55 kDa
CC   protein exists both in cells and in the medium, (2) that it
CC   undergoes oligomerisation, (3) phosphorylation, (4) glycosylation,
CC   (5) that it is susceptible to metalloprotease, (6) that it
CC   exhibits high affinity binding to extracellular matrix and HTM
CC   cells, (7) that it exhibits progressive inductions with time in
CC   both cell and organ cultures, and (8) that it exhibits high
CC   expression in the HTM of glaucomatous patients as compared to
CC   normal patients. TIFR* cDNA, the protein itself, molecules that
CC   bind it, and nucleic acid molecules that encode it, provide
CC   improved methods and reagents for diagnosing glaucoma and related
CC   disorders, such as cardiovascular and immunological diseases that
CC   affect expression of TIGR*. A claimed method of diagnosing glaucoma
CC   involves determining if the amount of TIGR* present in the HTM
CC   exceeds the amount found in an individual not predisposed to the
CC   disease.
SQ   Sequence   504 AA;


  Query Match             52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 TQKITYRISGVGID 15
          ||: |:||  || |
Db    281 TQETTWRIDTVGTD 294


RESULT   9
W73500
ID   W73500 standard; protein; 504 AA.
AC   W73500;
DT   26-FEB-1999  (first entry)
DE   Trabecular meshwork induced glucocorticoid response protein.
KW   TIGR protein; trabecular meshwork induced glucocorticoid response;
KW   secretory protein; antibody; glaucoma; diagnosis.
OS   Homo sapiens.
PN   US5849879-A.
PD   15-DEC-1998.
PF   14-MAY-1996; 645900.
PR   14-MAY-1996; US-645900.
PR   03-NOV-1994; US-336235.
PR   20-OCT-1995; US-546568.
PA   (REGC ) UNIV CALIFORNIA.
PI   Huang W, Nguyen TD, Polansky JR;
DR   WPI; 99-069807/06.
PT   Antibody to trabecular meshwork protein - useful for diagnosis of
PT   glaucoma
PS   Claim 1; Column 25-28; 22pp; English.
CC   This sequence represents the human Trabecular meshwork induced
CC   glucocorticoid response (TIGR) protein. The TIGR protein is a secretory
CC   protein specifically bound by the antibody of the invention. The
CC   antibody, especially in labelled form, can be used in the diagnosis of
CC   glaucoma by detecting elevated levels of the protein in the trabecular
CC   meshwork of the eye. Using the antibody, glaucoma is detected more
CC   accurately.
SQ   Sequence   504 AA;


  Query Match             52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 TQKITYRISGVGID 15
          ||: |:||  || |
Db    281 TQETTWRIDTVGTD 294


RESULT  10
W89391
ID   W89391 standard; Protein; 504 AA.
AC   W89391;
DT   08-MAR-1999  (first entry)
DE   Human trabecular meshwork induced glucocorticoid response protein.
KW   Human; trabecular meshwork induced glucocorticoid response protein;
KW   TIGR; glaucoma; primary open angle glaucoma; POAG; pigmentary glaucoma;
KW   low tension glaucoma; intraocular pressure; steroid; corticosteroid.
OS   Homo sapiens.
PN   US5854415-A.
PD   29-DEC-1998.
PF   25-JUN-1997; 882238.
PR   17-MAY-1996; US-649432.
PR   03-NOV-1994; US-336235.
PR   20-OCT-1995; US-546568.
PR   25-JUN-1997; US-882238.
PA   (REGC ) UNIV CALIFORNIA.
PI   Huang W, Nguyen TD, Polansky JR;
DR   WPI; 99-095006/08.
DR   N-PSDB; V81910.
PT   New isolated glaucoma-associated nucleic acids - which encode
PT   Trabecular Meshwork Induced Glucocorticoid Response protein, used to
PT   develop products for diagnosing glaucoma-related diseases
PS   Claim 1; Fig 1; 22pp; English.
CC   The present sequence is a human secretory protein from clone II.2.
CC   The secretory protein is designated TIGR (Trabecular Meshwork Induced
CC   Glucocorticoid Response) protein. The protein is highly induced by
CC   glucocorticoids in the endothelial lining cells of the human trabecular
CC   meshwork. The TIGR polynucleotides and proteins can be used as markers
CC   for the diagnosis of glaucoma, primary open angle glaucoma (POAG),
CC   pigmentary glaucoma, and low tension glaucoma and their related
CC   diseases. They can also be used to diagnose or protect an individual's
CC   sensitivity to elevated intraocular pressure upon administration of
CC   steroids such as glucocorticoids or corticosteroids. These products can
CC   also be used for diagnosing other diseases or conditions that affect the
CC   expression or activity of the protein. The products can also be
CC   formulated for administration to patients.
SQ   Sequence   504 AA;


  Query Match             52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 TQKITYRISGVGID 15
          ||: |:||  || |
Db    281 TQETTWRIDTVGTD 294


RESULT  11
R06441
ID   R06441 standard; protein; 309 AA.
AC   R06441;
DT   08-OCT-1990  (first entry)
DE   MG-6 antigen.
KW   Mycoplasma gallisepticum; poultry; vaccine.
PN   J02111795-A.
PD   24-APR-1990.
PF   02-JUN-1989; 136343.
PR   02-JUN-1989; JP-136343.
PA   (JAPG) Nippon Zeon KK, (SHIO) Shionogi KK.
DR   WPI; 90-169109/22.
DR   N-PSDB; Q05653.
PT   Diagnostic and vaccine for poultry mycoplasma serum - utilises
PT   antigen protein of the disease and recombinant vector
```

```
RESULT   6
W60670
ID   W60670 standard; Protein; 504 AA.
AC   W60670;
DT   14-SEP-1998  (first entry)
DE   Human glaucoma associated protein GLC1A.
KW   Glaucoma; GLC1A; treatment; mutant; juvenile open angle glaucoma;
KW   JOAG.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 178
FT                   /label= unknown
FT                   /note= "encoded by NCA"
PN   WO9820131-A1.
PD   14-MAY-1998.
PF   07-NOV-1997; U20702.
PR   21-MAR-1997; US-822999.
PR   08-NOV-1996; US-748479.
PR   30-JAN-1997; US-791347.
PA   (IOWA ) UNIV IOWA RES FOUND.
PI   Alward WLM, Sheffield V, Stone EM;
DR   WPI; 98-286947/25.
DR   N-PSDB; V37618, V37619.
PT   New isolated gene associated with glaucoma - used to develop
PT   products to determine whether a subject has, or is at risk of,
PT   developing glaucoma, and for treating or preventing glaucoma
PS   Disclosure; Pages 92-94; 116pp; English.
CC   This represents a human GLC1A protein sequence. The human GLC1A gene is
CC   associated with juvenile open angle glaucoma (JOAG). The gene can be used
CC   for the development of assays for identifying molecules that modulate
CC   (agonists or antagonists) the bioactivity of a functional or mutant gene
CC   or protein. Modulators may be an antibody, protein, peptide or
CC   peptidomimetic or a nucleic acid, e.g. antisense sequence, ribozyme or
CC   triple helix forming nucleic acid. These molecules can be administered to
CC   a subject with glaucoma or at risk for developing glaucoma to prevent or
CC   reduce the severity of the condition. Derivatives of GLC1A gene can be
CC   used to detect lesions of the GLC1A gene which are indicative of glaucoma
CC   or predisposition to glaucoma.
SQ   Sequence   504 AA;


  Query Match           52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 TQKITYRISGVGID 15
           ||: |:||  || |
Db     281 TQETTWRIDTVGTD 294


RESULT   7
W64669
ID   W64669 standard; Protein; 504 AA.
AC   W64669;
DT   27-OCT-1998  (first entry)
DE   Human TIGR protein.
KW   TIGR; trabecular meshwork induced glucocorticoid response protein; human;
KW   diagnosis; glaucoma; polymorphism; steroid sensitivity.
OS   Homo sapiens.
PN   WO9832850-A1.
PD   30-JUL-1998.
PF   09-JAN-1998; U00468.
PR   26-SEP-1997; US-938669.
PR   28-JAN-1997; US-791154.
PA   (REGC ) UNIV CALIFORNIA.
PI   Chen H, Chen P, Nguyen TD, Polansky JR;
DR   WPI; 98-427946/36.
DR   N-PSDB; V51391.
PT   Use of TIGR nucleic acid sequences - used for, e.g. developing
PT   products for diagnosis, prognosis and treatment of glaucoma
PS   Claim 48; Fig 8; 105pp; English.
CC   This sequence represents a novel human trabecular meshwork induced
CC   glucocorticoid response protein (TIGR) which is used in a method for
CC   diagnosing glaucoma in a patient. The method involves the detection of
CC   polymorphisms whose presence is predictive of a mutation affecting TIGR
CC   response in the patient and can be diagnostic of glaucoma or steroid
CC   sensitivity. Base substitutions and base additions upstream of and within
CC   TIGR exons can also be used to diagnose glaucoma.
SQ   Sequence   504 AA;


  Query Match           52.7%;  Score 39;  DB 1;  Length 504;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 TQKITYRISGVGID 15
           ||: |:||  || |
Db     281 TQETTWRIDTVGTD 294


RESULT   8
W70496
ID   W70496 standard; Protein; 504 AA.
AC   W70496;
DT   11-JAN-1999  (first entry)
DE   Trabecular meshwork induced glucocorticoid response* protein.
KW   Trabecular meshwork induced glucocorticoid response*; TIGR*;
KW   human; glaucoma; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1. .20
FT                   /label= Sig_peptide
FT   Protein         21. .504
FT                   /label= Mat_protein
FT   Modified_site   57. .60
FT                   /note= "Asn is N-glycosylated"
FT   Modified_site   231. .232
FT                   /note= "O-glycosylation"
FT   Modified_site   232. .233
FT                   /note= "O-glycosylation"
FT   Modified_site   277. .279
FT                   /note= "O-glycosylation"
FT   Modified_site   312. .313
FT                   /note= "O-glycosylation"
FT   Modified_site   404. .408
FT                   /note= "O-glycosylation"
FT   Modified_site   460. .464
FT                   /note= "O-glycosylation"
FT   Modified_site   464. .466
FT                   /note= "O-glycosylation"
FT   Binding_site    156. .160
FT                   /note= "heparin sulphate binding"
FT   Domain          233. .234
FT                   /note= "initiation domain"
FT   Domain          238. .239
FT                   /note= "initiation domain"
FT   Domain          331. .332
FT                   /note= "initiation domain"
FT   Region          85. .92
FT                   /note= "consensus leucine zipper unit"
FT   Region          92. .99
FT                   /note= "consensus leucine zipper unit"
FT   Region          131. .138
FT                   /note= "consensus leucine zipper unit"
FT   Region          138. .145
FT                   /note= "consensus leucine zipper unit"
FT   Region          145. .152
FT                   /note= "consensus leucine zipper unit"
FT   Region          152. .159
FT                   /note= "consensus leucine zipper unit"
FT   Region          159. .166
FT                   /note= "consensus leucine zipper unit"
PN   WO9844108-A1.
PD   08-OCT-1998.
PF   07-APR-1997; U05801.
PR   01-APR-1997; WO-U05391.
```